

## 深度学习在图像识别中的应用研究综述

郑远攀<sup>1,2</sup>, 李广阳<sup>1</sup>, 李 晔<sup>1</sup>

1. 郑州轻工业大学 计算机与通信工程学院, 郑州 450001

2. 应急平台信息技术河南省工程实验室, 郑州 450001

**摘 要:**深度学习作为图像识别领域重要的技术手段,有着广阔的应用前景,开展图像识别技术研究对推动计算机视觉及人工智能的发展具有重要的理论价值和现实意义,文中对深度学习在图像识别中的应用给予综述。介绍了深度学习的由来,具体分析了深度信念网络、卷积神经网络、循环神经网络、生成式对抗网络以及胶囊网络等深度学习模型,对各个深度学习模型的改进型模型逐一对比分析。总结近年来深度学习在人脸识别、医学图像识别、遥感图像分类等图像识别应用领域取得的研究成果并探讨了已有研究值得商榷之处,对深度学习在图像识别领域中的发展趋势进行探讨,指出有效使用迁移学习技术识别小样本数据,使用非监督与半监督学习对图像进行识别,如何对视频图像进行有效识别以及强化模型的理论性等是该领域研究的进一步方向。

**关键词:**深度学习;图像识别;卷积神经网络;胶囊网络;迁移学习;非监督学习

**文献标志码:**A **中图分类号:**TP391.41;TP181 **doi:**10.3778/j.issn.1002-8331.1903-0031

郑远攀,李广阳,李晔.深度学习在图像识别中的应用研究综述.计算机工程与应用,2019,55(12):20-36.

ZHENG Yuanpan, LI Guangyang, LI Ye. Survey of application of deep learning in image recognition. Computer Engineering and Applications, 2019, 55(12):20-36.

### Survey of Application of Deep Learning in Image Recognition

ZHENG Yuanpan<sup>1,2</sup>, LI Guangyang<sup>1</sup>, LI Ye<sup>1</sup>

1. School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450001, China

2. Henan Engineering Laboratory of Emergency Platform Information Technology, Zhengzhou 450001, China

**Abstract:** As an important technical means in the field of image recognition, deep learning has broad application prospects. Carrying out image recognition technology research has important theoretical and practical significance for promoting the development of computer vision and artificial intelligence. The application of deep learning in image recognition gives a review. The origin of deep learning is introduced. Deep learning models such as deep belief network, convolutional neural network, cyclic neural network, generated confrontation network and capsule network are analyzed. The improved models of each deep learning model are compared and analyzed one by one. In this paper, the research results of deep learning in image recognition applications such as face recognition, medical image recognition and remote sensing image classification are summarized. The existing researches are worth discussing. The development trend of deep learning in the field of image recognition is carried out. The discussion points out that the effective use of migration learning technology to identify small sample data, the use of unsupervised learning and semi-supervised learning to identify images, how to effectively identify video images and the theoretical significance of the model are further directions in this field.

**Key words:** deep learning; image recognition; convolutional neural network; capsule network; transfer learning; unsupervised learning

**基金项目:**国家自然科学基金(No.51404216);河南省高等学校青年骨干教师资助计划(No.2015GGJS-184);河南省科技攻关项目(No.152102310374);郑州轻工业大学青年骨干教师资助计划(No.2013XGGJS001);应急平台信息技术河南省工程实验室开放基金(No.YJ2013005);郑州轻工业大学校内科研基金(No.2015XJJY010)。

**作者简介:**郑远攀(1983—),男,博士,副教授,硕士生导师,研究方向为应急平台信息技术,E-mail:y.p.zheng@foxmail.com。

**收稿日期:**2019-03-04 **修回日期:**2019-04-18 **文章编号:**1002-8331(2019)12-0020-17

**CNKI网络出版:**2019-04-19, <http://kns.cnki.net/kcms/detail/11.2127.TP.20190418.1447.004.html>

## 1 引言

图像识别技术起源于20世纪40年代,当时由于技术不足以及硬件设施不完善,图像识别技术未得到快速发展。直到20世纪90年代,人工神经网络与支持向量机相结合,促进了图像识别技术的发展,图像识别技术得以广泛应用,例如在车牌识别、人脸识别、物体检测等方面。但是,传统的图像识别技术是以浅层次结构模型为主,需要人为对图像进行预处理,导致降低图像识别的准确率。针对此问题,科研工作者开始研究更深层次的网络结构模型,用模型自身提取图像特征,避免人为干预,在此期间,许多深度学习模型被提出,如:DBN、DFN、CNN、RNN等。深度学习的目的是通过构建一个多层网络,在此网络上计算机自动学习并得到数据隐含在内部的关系,提取出更高维、更抽象的数据,使学习到的特征更具有表达力。因此深度学习在图像识别中的应用研究是现在和未来很长一段时期内图像识别领域的重要研究课题,拟对本领域已有的研究进行调研并综述,为推动该方面的研究之进一步发展做出尝试。首先对深度学习的基本概念作一简介,其次对深度学习常用的结构模型进行概述说明,主要简述了深度信念网络(DBN)、卷积神经网络(CNN)、循环神经网络(RNN)、生成式对抗网络(GAN)、胶囊网络(CapsNet)以及对各个深度模型的改进模型做一对比分析。然后总结近几年深度学习在人脸识别、医学图像识别、遥感图像分类等应用领域取得的科研成果,并对在研究过程中的不足之处给予分析。最后对深度学习在图像识别领域的发展趋势进行探讨,具体阐述了图像识别未来发展趋势。

## 2 深度学习

### 2.1 深度学习方法和架构

深度学习按照学习架构可分为生成架构、判别架构及混合架构<sup>[1]</sup>。其生成架构模型主要包括:受限玻尔兹曼机<sup>[2]</sup>、自编码器<sup>[3]</sup>、深层信念网络<sup>[4]</sup>等。判别架构模型主要包括:深层前馈网络<sup>[5]</sup>、卷积神经网络<sup>[6]</sup>等。混合架构模型则是这两种架构的集合。

深度学习按数据是否具有标签可分为非监督学习与监督学习<sup>[7]</sup>。非监督学习方法主要包括:受限玻尔兹曼机、自动编码器、深层信念网络、深层玻尔兹曼机等。监督学习方法主要包括:深层感知器、深层前馈网络、卷积神经网络、深层堆叠网络、循环神经网络等。大量实验研究表明,监督学习与非监督学习之间无明确的界限,如:深度信念网络在训练过程中既用到监督学习方法又涉及非监督学习方法。

### 2.2 深度学习的发展

在机器学习中,许多学术界人士认为机器学习模型大致分为两类:浅层次结构模型和深层次结构模型<sup>[8]</sup>,对

应浅层学习和深度学习。随着浅层结构模型的不断完善,在基于浅层次结构模型的基础上,提出了解决一系列问题的算法模型,如:支持向量机(Support Vector Machines, SVM)<sup>[9]</sup>、Boosting<sup>[10]</sup>、最大熵方法(如, Logistic Regression, LR)<sup>[11]</sup>等。大量实验和实践验证,浅层次结构模型在处理图像、视频、语音、自然语言等高维数据方面表现较差,特征提取难以满足需求,而深度学习技术弥补了这一缺陷,在提取物体深层次的结构特征方面更具有优势。

2006年, Hinton<sup>[12]</sup>等人提出深度学习的概念,它是含多隐藏层、多感知器的一种网络结构,能更抽象、更深层次描述物体的属性和特征。由于深度学习算法是高维矩阵运算,之前由于计算机性能的缺陷,运算能力弱,运算速度慢,无法完成大规模运算,深度学习的发展遇到了瓶颈<sup>[13]</sup>。随着计算机性能的不不断提升,针对高性能运算的硬件不断完善,在一定程度上提高了计算机的运算能力和运算速度,现有高性能计算机可以完成深度学习中大规模的矩阵运算。因此,深度学习在这段时间内得到了快速发展,针对深度学习的各种算法模型不断提出,并且顺利运用到商业之中。例如:在国内,以百度为代表的无人驾驶汽车研发成功,其中将以深度学习为基础的计算机视觉、听觉等识别技术应用到“百度无人驾驶汽车”系统研发中,通过一系列的测试,最终,百度无人车亮相央视春晚,并且在珠港澳大桥上完成无人驾驶首跑。另一方面以阿里巴巴为代表的图像搜索识别技术取得了很好的成绩,在此技术之上研制出一系列的商业产品,如:基于图像搜索的“拍立淘”等。除此之外,商汤科技在智能视频、身份验证、移动互联网等方面全部运用大量的深度学习技术。在国外, Google、Microsoft、Facebook 等公司将深度学习技术作为公司发展的重要技术之一,甚至针对深度学习技术研发了一系列的深度学习框架,如: Google公司研发的Tensorflow, Microsoft公司研发的CNTK, Facebook公司研发的Torch, Fchollet公司研发的Keras, DMLC研发的MXNet以及BLVC (Berkeley Vision and Learning Center)和社区贡献者共同研发的Caffe等,这些深度技术框架主要应用于图像识别分类、手写字识别、语音识别、预测、自然语言处理等方面。正是由于上述深度学习框架的提出与应用,促进了深度学习技术的快速发展。其主流深度学习框架的对比分析如表1所示。

## 3 图像识别中的深度学习模型

### 3.1 深层信念网络

深层信念网络(Deep Belief Network, DBN)起源于人工神经网络,是一个概率生成模型,由多层受限玻尔兹曼机(RBM)和一层某种分类器组合而成,经典的DBN网络结构是由若干层RBM和一层BP组成的一种

表1 主流深度学习框架对比分析

框架	主要开发语言	应用场景	优点	缺点
Tensorflow	Python/C++/Go	图像识别, 手写字识别, 语音识别, 预测, 自然语言处理	兼容性好, 易学, 易扩展, 支持并行运算 提供了可视化工具 支持多种编程语言以及支持细粒度网络层, 方便调试复杂的计算图模型	系统设计过于复杂 语言接口变动频繁 计算复杂, 执行效率较低 缺少对视频识别有用的三维卷积
CNTK	C++	图像识别, 手写字识别, 语音识别	兼容性好 支持细粒度网络层 支持跨平台运行 计算性能强	不支持 ARM 架构, 限制了其在移动设备上的功能
Torch	Lua/Python	图像识别, 手写字识别, 语音识别	兼容性好, 运行速度快, 易学易用 设计简洁, 后期维护强大	编程语言主要为 Lua, Lua 不是主流语言, 在开发人员没有熟练掌握 Lua 之前, 使用 Torch 很难提高开发的整体生产力 缺乏多种编程语言接口
Keras	Python	图像识别, 手写字识别, 语音识别	易学, 支持快速实验	过度封装导致丧失灵活性, 程序运行速度慢
MXNet	R/Julia/Python/C++	图像识别, 手写字识别, 语音识别, 预测, 自然语言处理	兼容性、扩展性以及移植性最强 支持混合编程和多种编程语言接口 支持生成对抗网络(GAN)模型	接口文档混乱
Caffe	C++/Python	图像识别, 视频图像识别	运行速度快 支持跨平台运行 支持多种编程语言接口	不支持细粒度网络层, 构建复杂的层类型必须以低级语言完成 对循环网络的支持上很薄弱

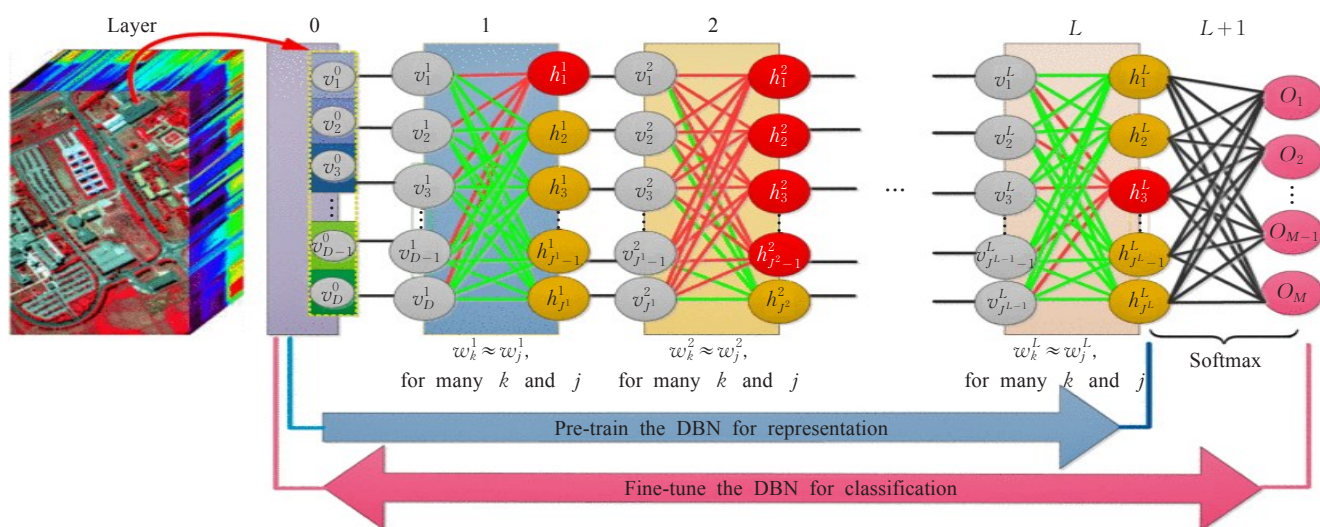


图1 高光谱图像的DBN模型图

深层神经网络。此深度模型广泛应用于图像分类识别, 语音识别等领域<sup>[14]</sup>。

2017年, Zhong 等人<sup>[15]</sup>提出一种新型的DBN对高光谱遥感图像进行分类, 但是通常的预训练和微调方法会使学习DBN中的许多隐藏单元表现得非常相似, 或者表现为“从不响应”或“总是响应”。这些结果会对DBN的描述能力和分类性能产生负面影响。为了进一步提高DBN的性能, 作者通过对预训练和微调程序进行规范化, 以多样性提升先验因素和潜在因素, 开发了一种新的多样化DBN。此外, 通过常用的递归贪婪和反向传播学习框架, 可以有效地实现规则化的预训练和微调。其高光谱图像的DBN的图形表示如图1所示。

由图1所示,  $D$  维输入  $(v_1^0, v_2^0, \dots, v_D^0)^T$  可以是像素

的光谱特征, 或者是在像素及其相邻点上提取的特征。然后对RBM序列进行分层叠加, 使用前一个RBM的输出作为当前RBM的输入。在DBN中, 两个相邻的层之间有一组完整的连接, 但同一层中没有两个单元被连接。DBN的结构表明, 第一个隐藏层将输入的光谱信息合并, 随后一个或多个隐藏层将详细学习光谱特征。每层都可以输出输入数据的一种表示或特征, 层越高, 特征越抽象。由于最后的任务是对高光谱图像进行分类, 因此添加了一个SoftMax层作为DBN的最后一层, 对从最后一个RBM中提取的特征进行分类。主要过程包括预训练和微调。

预训练学习阶段, 作者所提出的训练方法是通过一个递归贪婪学习过程来实现的, 该过程使权重参数多样



化,并间接地使相应的潜在单位多样化。多样化方法可以概括为以下步骤:(1)用初始输入训练第一层RBM,使参数多样化,使之适合训练数据;(2)用学习的多样化参数 $w^1$ 计算多样化隐藏 $h^1$ ;(3)用隐藏 $h^1$ 作为输入数据训练下一层RBM;(4)重复程序直到所有随后的隐藏层 $h^l$ 在 $l=1,2,\dots,L$ 时都是多样化,其中 $l$ 是DBN中隐藏层的数量。

在预训练学习阶段,作者使用多样化参数训练出多样化的隐藏层,一方面使参数及隐藏层具有更丰富的表达,进一步提取图像的深层次的特征。另一方面使“从不响应”及“总是响应”的神经元得到良好的调节,尽可能使每个神经元处于正常状态。

微调阶段,采用监督学习方法进行多样化学习,BP算法将高光谱图像分类的主要目标语义标签的信息传播回以前的所有层,使参数最适合训练数据。同时,通过BP算法实现的分散化方法可以对不同层次参数进行联合学习和分散化。然后,在微调过程中采用多样化的方法,有可能充分利用它们的相关性,并可能进一步提高DBN的表示能力。

在微调阶段,采用监督学习方法对参数进行调节,这是由于无监督的方式将所有层训练在一起通常非常困难,通过在有监督的微调过程中执行多样化的方法来解决这个问题,该过程将分类错误定义在标记的训练样本上传播到整个网络,从而使用语义信息来调整所有参数。

对高光谱图像的实验表明,通过预训练和微调过程中的多样性正则化,学习到的DBN具有更多的潜在因素,这直接使得多样化的DBN比原始的DBN获得更好的效果,并与其他最新的高光谱图像分类方法相比较具

有更好的性能。

由此可分析出,作者使用多样化的DBN进行遥感图像分类虽然在图像分类精度、时间及各项性能上取得了较好的效果,限于DBN的局限性,对图像深层次的特征较弱,参数量较大,图像依然会出现一定程度的过拟合现象。但是作者提出的解决方案为DBN在图像分类中的应用具有重大的参考及研究价值。

### 3.2 卷积神经网络

卷积神经网络(Convolutional Neural Network,CNN)是一种特殊的深层前馈网络,CNN模型主要包含输入层、卷积层、池化层、全连接层以及输出层。但是,在网络结构中,为了使输出更加准确,特征提取更加丰富,通常网络模型中使用多卷积层和多池化层相结合的网络模型,较为经典的CNN模型有LeNet-5、AlexNet、ZF-Net、VGGNet、GoogLeNet、ResNet以及DenseNet<sup>[16]</sup>,上述CNN模型均是LeNet的改进型模型。

2019年,Zhang等人<sup>[17]</sup>提出一种多特征的加权DenseNet(MFR-DenseNet)用于图像分类。MFR-DenseNet通过自适应地重新校准信道特征响应和明确地建模不同卷积层特征之间的相互依赖性,提高了DenseNet的表示能力。首先,为了进行动态信道特征重新校准,作者将挤压激励模块(SEM)引入到DenseNet中,构造了信道特征Reweight DenseNet(CFR-DenseNet)。然后,为了模拟不同卷积层特征之间的相互依赖关系,提出了双挤压激励模块(DSEM),并构造了层间特征Reweight Densenet(ILFR-DenseNet)。最后,将CFR-DenseNet和ILFR-DenseNet与集成学习方法结合起来设计了MFR-DenseNet。其模型结构如图2所示。

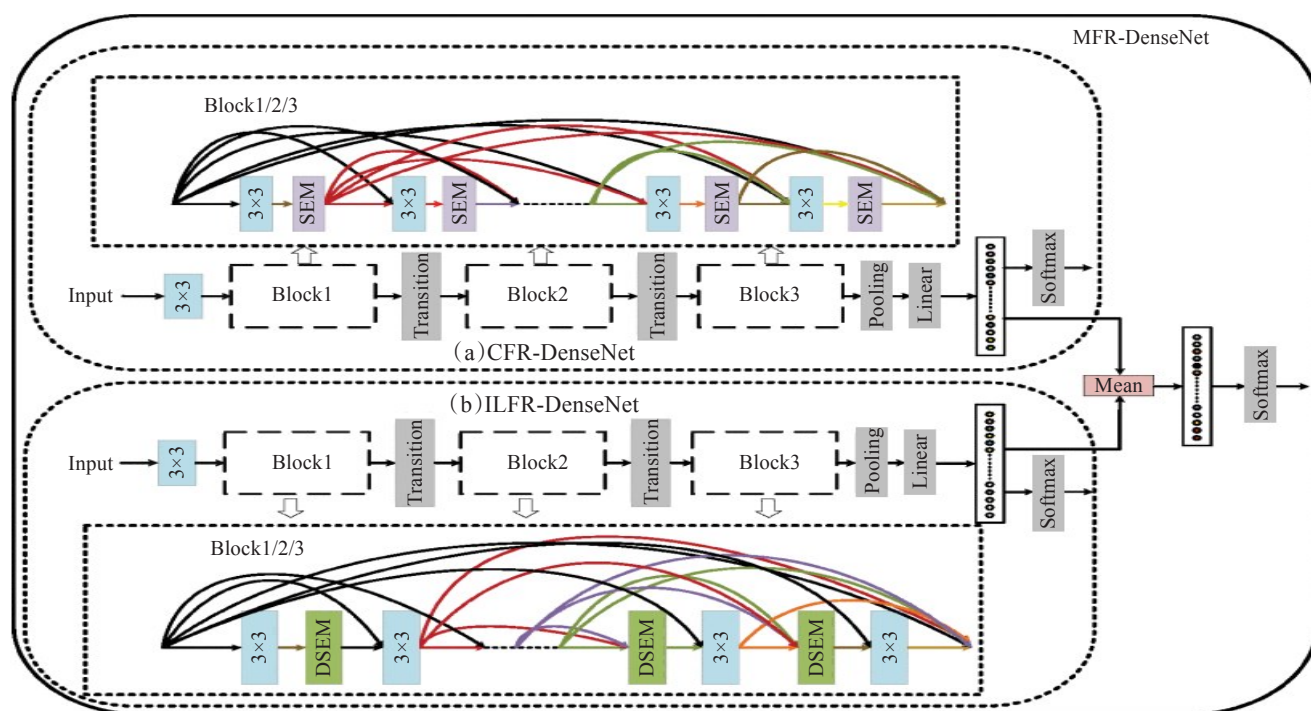


图2 MFR-DenseNet结构模型图

(1)CFR-DenseNet: DenseNet 中每个卷积层的输入只是由前面所有层生成的所有特征图的串联。作者通过引入 SEM 建立了 CFR-DenseNet 模型,该模型模拟了特征通道之间的相互依赖性。图 2(a)显示了 CFR-DenseNet 架构中的一个密集块。每  $3 \times 3$  卷积层后添加了 SEM。网络通过自动学习获得每个特征信道的权重,然后根据权重增强有用的特征,从而抑制对当前任务不有用的特征。该体系结构明确地对卷积层的信道特征重新校准建模。

(2)ILFR-DenseNet: DenseNet 最明显的特点是卷积层的输入是所有前面层的输出,而 Vggs 和 ResNet 等结构是由许多卷积层叠加而成的。针对这一特殊性,作者提出了 DSEM,它明确地模拟了层间特征的相互依赖性。通过在  $3 \times 3$  卷积层之前添加 DSEM 来构建 ILFR-DenseNet,如图 2(b)所示。

(3)MFR-DenseNet: 为了同时探索信道之间的相关性和层间特征的相互依赖性,构建了基于集成学习的融合了 CFR-DenseNet 和 ILFR-DenseNet 的 MFR-DenseNet。MFR-DenseNet 最大限度地发挥了 CFR-DenseNet 和 ILFR-DenseNet 的优势。首先,培训并保存 CFR-DenseNet 和 ILFR-DenseNet 模型。其次,加载模型和权重。在测试中,从两个模型中取平均预测值,并使用它进行最终预测(SoftMax)。

然而,MFR-DenseNet 模型结构中需要解决以下三个问题,才能使模型达到最优。① 如何将 SEM 集成到 DenseNet 中? ② 用什么方法训练 ILFR-DenseNet? ③ 哪一块或层需要添加 DSEM?

针对这些问题,作者通过实验给出了较好的解决方案,① 在每个卷积层之后向 DenseNet 添加 SEM,然后将输出与前面层的特征图连接起来。② 在训练方法中,需要分三步完成。首先,使用端到端的方法来训练 CFR-DenseNet 并保存最佳模型和权重。其次,将第一步中保存的激励操作的权重加载到 ILFR-DenseNet 中相应卷积层的第一激励操作中,并且在不进行训练的情况下固定这些权重。在最后,除了加载的权重外,还初始化了模型中的权重,并使用端到端的方法来训练 ILFR-DenseNet。③ 将 DSEM 添加到每个块的最后一层。

DenseNet 是 CNN 改进模型中技术最新的图像分类识别模型之一,在图像分类识别方面,由于 DenseNet 及其他改进的 CNN 在结构上具有其他神经网络不具备的优势,正是由于这些结构特征,使卷积神经网络在图像识别和物体检测方面取得了巨大的成功<sup>[18]</sup>。其改进型 CNN 模型的对比分析如表 2 所示。

### 3.3 循环神经网络

循环神经网络(Recurrent Neural Network, RNN),

表 2 改进型 CNN 模型的对比分析

模型	技术手段	结构特点
LeNet-5	激活函数为 Relu, 分类器使用的 Softmax 回归	网络结构简单, 模型深度较浅, 图像特征提取能力一般 训练过程中容易出现过拟合
AlexNet	激活函数为 Relu, 采用 dropout 技术, 数据增强技术, 多 GPU 平行训练技术等	有效避免过拟合现象 网络模型的收敛速度会相对稳定, 能避免或抑制网络训练时的梯度消失现象 模型训练速度较快。具有更深的网络结构, 计算量增大, 具有更多的参数
ZF-Net	激活函数为 Relu, 采用 dropout 技术, 数据增强技术, 多 GPU 平行训练技术, 使用较小 filter, 分类器使用的 Softmax 回归等	调节了参数, 性能比 AlexNet 更强, 保留更多原始像素信息 网络结构没什么改进, 同 AlexNet
VGGNet	激活函数为 Relu, 采用 dropout 技术, 数据增强技术, 多 GPU 平行训练技术, 使用 $1 \times 1$ 和 $3 \times 3$ 的小卷积核, 分类器使用的 Softmax 回归等	小卷积核使判决函数更具有判决性, 具有更少的参数 增加了非线性表达能力, 网络结构更深, 计算量更大
GoogLeNet	激活函数为 Relu, 采用 dropout 技术, 数据增强技术, 多 GPU 平行训练技术, 引入 Inception 结构代替了单纯的卷积+激活的传统操作技术, 分类器使用的 Softmax 回归等	引入 Inception 结构, 使用 $1 \times 1$ 卷积核来降维, 解决计算量大的问题 中间层使用 LOSS 单元作为辅助更新网络 全连接层全部替换为简单的全局平均 pooling, 参数更少, 虽然网络深, 但参数只是 AlexNet 的 1/12 倍
ResNet	激活函数为 Relu, 多 GPU 平行训练技术, 引入残差块, 平均池化, 分类器使用的 Softmax 回归等技术	引入残差单元, 通过直接将输入信息绕道传到输出, 保护信息的完整性 整个网络只需要学习输入、输出差别的那一部分, 简化学习目标和难度 在一定程度上解决了信息传递的时候或多或少会存在信息丢失, 损耗等问题, 同时还有导致梯度消失或者梯度爆炸, 导致很深的网络无法训练等问题
DenseNet	激活函数为 Relu, 多 GPU 平行训练技术, 平均池化, 分类器使用的 Softmax 回归等技术	由若干个 Dense Block 串联起来而得到的, 在每个 Dense block 之间有一个 Convolution+Pooling 的操作, DenseNet 通过连接操作来结合 feature map, 并且每一层都与其他层有关系, 这种方式使得信息流最大化。解决了深层网络的梯度消失问题, 加强了特征的传播, 鼓励特征重用, 减少了模型参数



又名时间递归神经网络<sup>[19]</sup>,主要是用来解决序列数据问题。在RNN结构模型中,网络会对之前时刻的信息进行记忆并且运用到当前的输出计算之中,相比于卷积神经网络、深度前馈网络,循环神经网络隐藏层之间的神经元是相互连接的,隐藏层中神经元的输入是由输入层的输出和上一时刻隐藏层神经元的输出共同组成。

循环神经网络(RNN)是最早应用于语音识别<sup>[20]</sup>和手写字识别<sup>[21]</sup>,并且取得了显著效果,但是RNN在实际应用中仍有不足之处,如:训练难度大,效率低,时间长,准确度低等,随后,科研工作者在RNN模型的基础之上对RNN进行了改进,在一定程度上弥补了上述的不足,例如:长短时记忆网络(Long Short-Term Memory, LSTM)<sup>[22]</sup>,GRU(Gated Recurrent Unit)<sup>[23]</sup>,双向RNN(bidirectional RNN)<sup>[24]</sup>等模型。而这些改进的RNN模型逐步在图像分类识别方面表现出良好的效果。

2017年Mou等人<sup>[25]</sup>提出一种新的RNN模型,该模型可以有效地将高光谱像素作为序列数据进行分析,然后通过网络推理确定信息类别。具体地说,RNN利用一个新提出的激活函数-参数校正tanh(PRetanh),而不是流行的tanh或校正线性单元,用于高光谱序列数据分析。作者所提出的激活函数使得在训练过程中具有相当高的学习率,不存在具有分歧的风险。此外,为了有效地处理高光谱数据,减少参数总数,采用了一种改进的门控递归单元,该单元使用了PRetanh来进行隐式表示。其模型结构如图3所示。

由图3可知,作者提出的RNN模型流程图大致分为以下四个步骤:首先,将现有光谱带 $x^k$ 的值输入输入层。然后,循环层接收 $x^k$ 并计算当前波段的隐藏状态信息;同时恢复该信息。随后,将下一个波段 $x^{k+1}$ 的值与 $x^k$ 的状态信息同时输入到循环层,并通过激活和上一个波段 $k$ 的激活之间的线性插值计算光谱波段 $k+1$ 的激活。

最后,RNN通过循环遍历整个高光谱像素序列来预测输入高光谱像素的标签。

在新模型中,作者使用一种新的激活函数-参数校正tanh(PRetanh),Pretanh激活函数相比较于传统的激活函数(tanh, sigmoid, relu)具有以下两个优点:(1)产生有界输出。(2)自适应地提高稀疏度。

另一方面,作者使用一种改进的GRU以减少参数总数。由于GRU是LSTM的一种变形网络,具有比LSTM更简单的网络结构和更少的参数,因此,作者在GRU的基础之上添加了一个参数修正线性函数——PreLU,用来控制在光谱带上将要暴露的部分状态信息,而不是在每个光谱带上暴露整个状态值。

正由于RNN模型的不断改进,使RNN在高图像分类识别方面取得了一定的进展,为图像分类识别提供一种新的解决方案<sup>[26]</sup>。虽然,改进的RNN依然存在缺点,但不影响科研工作对RNN的进一步研究,将来将会有更新的RNN模型被提出。其改进型RNN对比分析如表3所示。

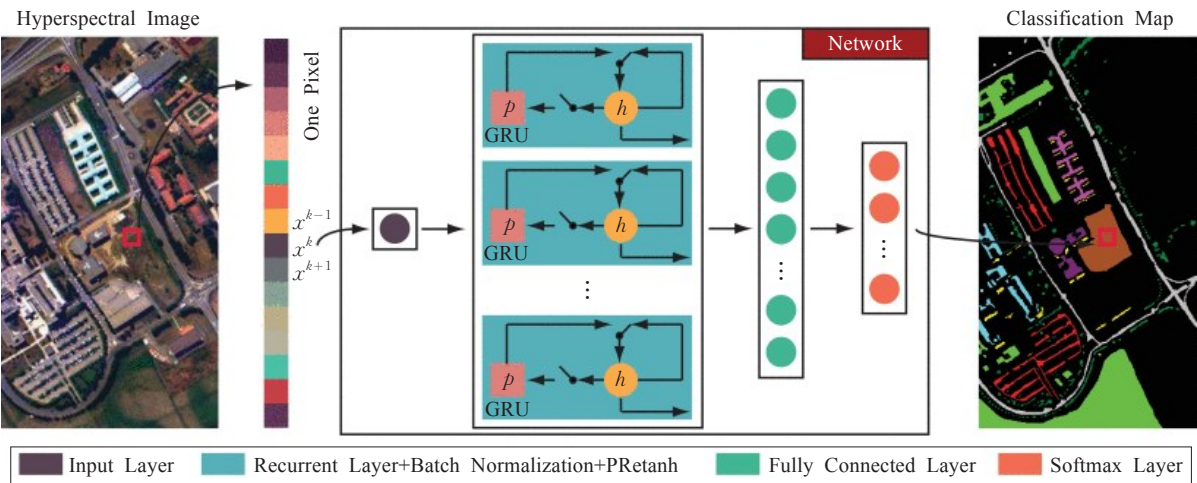


图3 高光谱像素RNN模型图

表3 改进型RNN对比分析

模型	应用场景	优点	缺点
LSTM	语音识别、图像描述、自然语言处理	解决了RNN存在着梯度消失或梯度爆炸等问题,能够学习长期依赖关系	网络结构很复杂,训练时间较长
GRU	语音识别、图像描述、自然语言处理	解决了RNN存在着梯度消失或梯度爆炸等问题,能够学习长期依赖关系,是LSTM的一种变形,结构比LSTM简单,具有更少的参数	训练时间虽然比LSTM有缩短,但实际时间仍较长
双向RNN	语音识别、图像描述、自然语言处理	模型从前向后保留该词前面的词的重要信息,同时从后向前去保留该词后面的词的重要信息	存在着梯度消失或梯度爆炸等问题

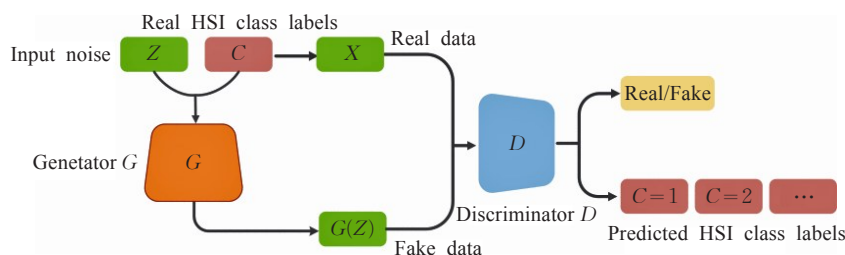


图4 生成式对抗网络结构模型图

### 3.4 生成式对抗网络

生成式对抗网络(Generative Adversarial Network, GAN)<sup>[27]</sup>是 Goodfellow 等人于 2014 年提出的一种生成式模型,通过在对抗过程中估计并生成模型的新框架,是近几年最成功的生成模型。GAN 主要由两部分构成:生成模型(G)和判别模型(D)。生成模型捕捉真实数据样本的潜在分布,并生成新的数据样本。判别模型是一个二分类器,判别区分输入的是真实数据还是生成的样本数据。判别模型输出是以概率值表示,概率值大于 0.5 则为真,概率值小于 0.5 则为假。当判别器无法区别出真实数据和生成数据时则停止训练,此时达到生成器与判别器之间判定误差的平衡,训练达到理想状态。

2018 年, Zhu 等人<sup>[28]</sup>首次探讨了 GAN 在高光谱数据分类中的应用。在 GANS 的帮助下,深度 CNN 在分类精度上取得了较好的效果, CNN 提出的过拟合问题得到了缓解。作者等人提出了两种用于 hyperspectral images(HIS)分类的框架: 1D-GAN 和 3D-GAN, 两种框架的分类结果表明,即使在训练样本有限的情况下,作者所提的 GAN 仍优于传统的 CNN。

在所提出的 GAN 中,设计了一个卷积神经网络(CNN)来区分输入,并使用另一个 CNN 来生成所谓的假输入。上述 CNN 是一起训练的:生成 CNN 试图生成尽可能真实的假输入,而识别 CNN 则试图对真实输入和假输入进行分类。这种对抗训练提高了识别 CNN 的泛化能力,在训练样本有限的情况下,这一点非常重要。

提出了两种用于 HIS 分类的框架: 1D-GAN 和 3D-GAN。1D-GAN 主要由  $G$  接收噪声  $Z$  和标签  $C$  作为输入,而  $D$  接收带有标签  $C$  的实际光谱矢量  $X$  和一些批次的假光谱矢量  $G(Z)$ 。然后,通过  $D$  网络,可以得到最终的结果,包括类标签和源标签。而 3D-GAN 是由发生器接受噪声作为输入,并将其转换为与真实数据相同的大小。然后,鉴别器接受真实数据或生成的假样本作为输入数据,并使用识别 CNN 分类器给出真实和假分类结果,使用 SoftMax 分类器给出分类图。在该框架中,使用两种 CNN,其两者的共同点是池层被跨步卷积(在识别器中)和分数跨步卷积(在生成器中)取代。此外,在生成器和识别器中都使用了批处理规范化,并且删除了与完全连接的隐藏层。其共同的模型结构如图 4 所示。

GAN 在图像生成方面取得了一定的进展,于是,一部分学者对 GAN 进行改进,将 GAN 逐步应用在图像分类识别等方面,如,商显震等人<sup>[29]</sup>提出一种融合生成对抗网络和朴素贝叶斯的多分类诊断方法,该方法使皮肤病诊断方法的准确率和召回率均有提升。唐贤伦等人<sup>[30]</sup>提出了一种结合深度卷积生成对抗网络(DCGAN)和条件生成对抗网络(CGAN)的优点,建立条件深度卷积生成对抗网络模型(Conditional-DCGAN, C-DCGAN),模型在 MNIST 数据集和 CIFAR-10 数据集上算法的识别正确率分别为 99.45% 和 84%。

但是 GAN 的缺点还是比较明显的,例如:针对 GAN 本身不可控的缺点, Mehdi 等人<sup>[31]</sup>提出了条件生成式对抗网络(CGAN)。为了能够产生更高的分辨率图像, Denton 等人<sup>[32]</sup>在 GAN 和 CGAN 的基础上,结合  $K$  层的拉普拉斯金字塔框架构成拉普拉斯对抗神经网络(LAPGAN)。为解决训练和收敛函数不稳定的问题, Arjovsky 等人<sup>[33]</sup>提出了 Wasserstein GAN 模型,但是 WGAN 并没有完全解决不稳定问题,此后 Gulrajani 等人<sup>[34]</sup>提出了 WGAN-GP 模型,有效地解决了模型不稳定问题等。文中列举了改进的 GAN 模型,其改进型 GAN 的改进特点及模型效果如表 4 所示。

### 3.5 胶囊网络

胶囊网络(Capsule Network, CapsNet)<sup>[38]</sup>是 Hinton 等人在 2017 年提出,是当前图像分类识别最前沿的技术之一。CapsNet 是在 CNN 的基础之上发展而来,解决了 CNN 对物体之间的空间辨识度差及物体大幅度旋转之后识别能力低下的两个缺陷。目前的 CapsNet 结构较浅,是由卷积层、PrimaryCaps(主胶囊)层、DigitCaps(数字胶囊)层构成。其 CapsNet 结构如图 5 所示(以一张  $28 \times 28$  的 MNIST 数字图像为例)。

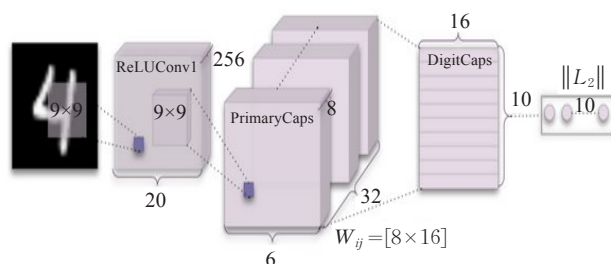


图5 胶囊网络结构模型图



表4 改进型GAN的主要改进特点及模型效果

模型	主要改进	模型效果
CGAN	生成器和判别器均增加了额外信息为条件	解决了GAN训练自由、模型不可控问题
DCGAN <sup>[35]</sup>	使用卷积层代替全连接层,判别器几乎和生成器是对称的 整个网络没有pooling层和上采样层,使用了带步长的卷积代替了上采样	增加的训练的稳定性
LAPGAN	卷积层使用高斯金字塔,用于下采样,使用拉普拉斯金字塔,用于上采样	用于生成高像素的图像
LSGAN <sup>[36]</sup>	GAN的目标函数由交叉熵损失换成最小二乘损失	提高生成图像质量,训练更稳定
WGAN	判别器最后一层去掉sigmoid,生成器和判别器的loss不取log 对更新后的权重强制截断到一定范围内,使用SGD,RMSprop等优化器	理论上解决了训练不稳定的问题 解决了模式崩溃的问题,生成结果多样性更丰富
WGAN-GP	改进了WGAN连续性限制的条件,提出了一种新的lipschitz连续性限制手法—梯度惩罚	解决了模型不稳定问题,训练梯度消失梯度爆炸的问题,比标准WGAN拥有更快的收敛速度,并能生成更高质量的样本
BEGAN <sup>[37]</sup>	提出了一种新的简单强大GAN,使用标准的训练方式,对于GAN中G、D的能力的平衡提出了一种均衡的概念 提出一种收敛程度的估计算法	训练速度快,收敛稳定

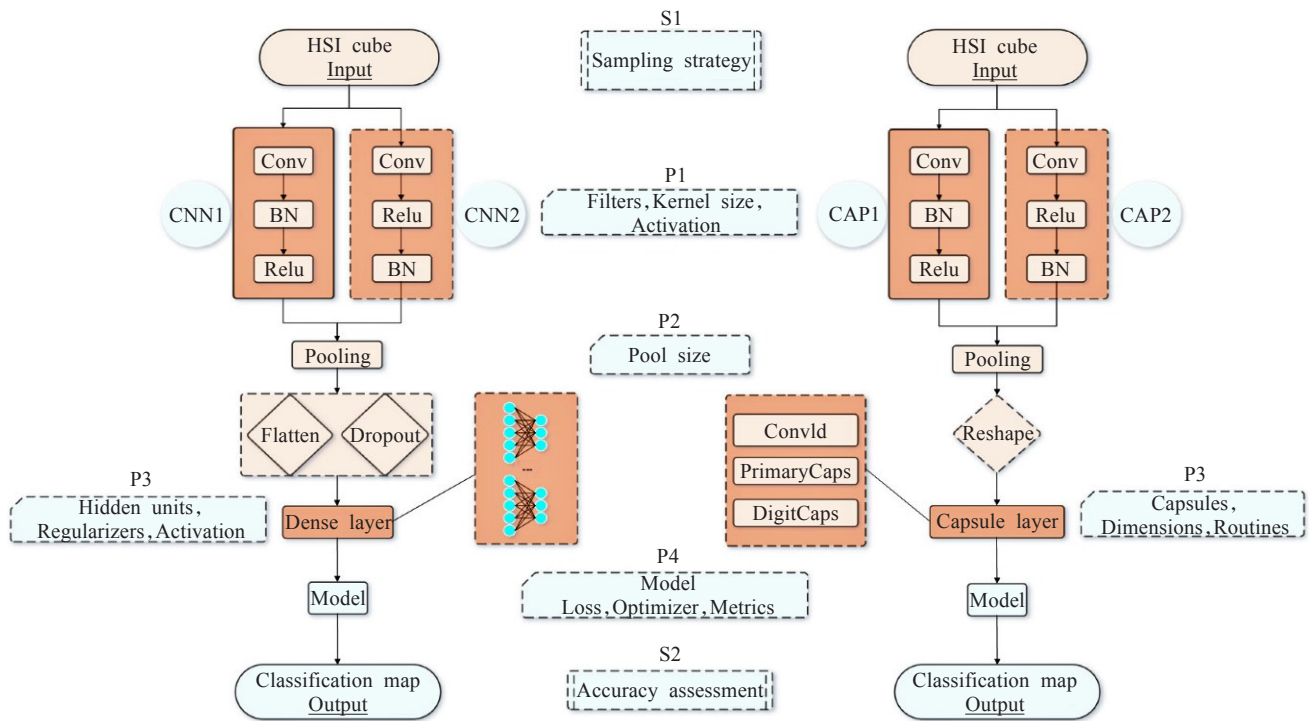


图6 CNN与CapsNet结构模型对比图

传统的神经网络结构是以神经元为载体,一个神经元表示物体的一个属性,CapsNet中的胶囊是由多个神经元共同组成,而CapsNet由多个胶囊组成,一方面丰富的胶囊的内部特征,另一方面给与神经网络更多的属性信息。

由于传统神经网络的BP训练算法不适用于CapsNet的训练,于是Sabour等人提出了胶囊之间新的训练算法——动态路由(dynamic routing)算法,该算法的主要思想可概况为:低级别胶囊需要将输出传递给允许该输出的高级别胶囊。具体就是,路由的数据传输是由低层次胶囊的输入(预测向量)与高层次胶囊的输出(激活向量)的相似度决定。如果低层胶囊的预测向量与高层胶囊的激活向量具有较高相似度,则判定两个胶囊是高度相关的。

2018年,Deng等人<sup>[39]</sup>提出基于有限训练样本的胶囊网络高光谱图像分类。提出了一种改进的两层有限训练样本CapsNet用于HSI分类。使用两个真实的HSI数据集(即Paviau(PU)和Salinasa数据集)对所提出的CapsNet进行训练,分别表示复杂和简单的数据集,并用于研究每个模型或分类器的鲁棒性或表示,此外,为了比较CNN和CapsNet,还提出了一种可比较的网络架构设计范例。其CNN和CapsNet结构模型设计如图6所示。

由图6可知,所有模型都是两层网络结构CNN1(或CAP1)与CNN2(或CAP2),前者包含一个复合块“Conv、BN、Relu”,后者包含一个传感器将等量的堆叠层与批处理规范化(BN)层和Relu结合在一起,使图层反转。两种模型最大的不同在于CNN的Dense Layer和CapsNet的Capsule Layer,前者主要由隐藏单元、正则化矩阵、



激活函数等组成,后者主要由由卷积ID、PrimaryCaps层、DigitCaps层构成。由于CapsNet结构的特殊性及其独特的训练算法,在实验过程证实,CapsNet较CNN具有更好的分类精度。

截止目前,胶囊网络在MNIST数据集分类任务中取得最好的成绩,由于胶囊网络相比于传统的CNN是一个浅层次网络模型,仅仅只有三层网络,但是错误识别率达到了0.25%,优于同级别的神经网络模型。由于胶囊网络的层次较浅,在图像分类识别方面落后于目前流行的CNN模型。将来在增加网络结构,优化路由算法等方面进一步研究,胶囊网络在图像分类识别中将会有很大的发展空间。

### 3.6 深度学习模型对比分析

由于深度学习模型存在各自优点与缺点,为了使读者对所列举的典型网络有更深刻的理解,文中其进行了对比分析,如表5所示。

## 4 深度学习在图像识别中的应用

深度学习因提取特征能力强,识别精度高,实时性快等优点被广泛应用。特别在人脸识别、医学图像识别、遥感图像识别等方面取得了不俗的效果。在深度学

习技术发展之前,图像识别以传统的技术手段实现,例如,主成分分析法(PCA)、拉普拉斯特征图法、局部保值映射法(LPP)、稀疏表示法、神经网络降维法等。但是,由于传统技术手段的缺陷,图像识别效果并不理想。随着深度学习技术的发展,科研工作者开始将深度学习应用到图像识别中,并且在最近的几年中取得了巨大的进步,图像的识别效率及识别效果比传统的识别手段有了较大的进步。其传统识别方法与深度学习识别方法对比分析如表6所示。

### 4.1 人脸识别

人脸识别技术是日常生活中最常用的技术手段,广泛应用在视频监控、门禁系统、电子商务等方面。构建人脸识别系统需要一定的技术手段,包括人脸图像采集,人脸定位,人脸识别预处理,身份验证及身份查找等。在实际研发过程中,人脸识别系统功能模块主要包括人脸捕捉及跟踪,人脸识别对比,人脸建模与检索,真人鉴别及图像质量检测等模块。其中,人脸识别对比是人脸识别系统中最具核心的技术之一,人脸识别对比的正确率直接影响着整个系统的精确度。在深度学习技术诞生之前,人脸识别对比算法以PCA、LDA、LBP及其改进算法为主,由于传统识别算法的局限性,人脸识别

表5 经典网络结构对比分析

方法	主要任务	优点	缺点
DBN	特征提取	能够反映同类数据本身的相似性	对于分类问题,分类精度不高 某些学习的复杂性较高 输入数据具有平移不变性
CNN	特征提取	训练参数减少,模型的泛化能力更强 池化运算降低网络的空间维度,对输入数据的平移不变性要求不高	容易出现梯度消散问题 空间关系辨识度差 物体大幅度旋转之后识别能力低下
RNN	特征提取	可以对序列内容建模	需要训练的参数较多,容易出现梯度消散或梯度爆炸问题 不具备特征学习能力
GAN	生成对抗样本	与其他模型相比,能够产生更好的样本 可以训练任何一种生成器网络 不需要设计遵循任何种类的因式分解的模型,任何生成器网络 and 任何鉴别器都会有用 避免了复杂的马尔科夫过程,回避了近似计算棘手的概率难题	模型的收敛性较差 训练过程中,容易出现崩溃问题 由于无需预先建模,模型过于自由不可控
CapsNet	特征提取	解决了CNN模型出现的空间关系辨识度差以及物体大幅度旋转之后识别能力低下	模型的网络结构较浅,在图像识别分类上的正确率和目前流行的CNN模型还是有很大差距

表6 图像识别方法对比分析

方法	优点	缺点
传统识别方法	模型结构层次较浅,计算量较小,计算时间较短 不需要以大量的图像为基础,即可完成图像的识别分析	无法从原图像中获取更高层次的语义特征和深度特征 由于人为以及外界因素的干扰,图像识别率较低 面对大数据的情况下,无人为设计,无法获取图像特征 需要以大量的图像为基础,目前图像数据集范围较窄,许多图像还未建立相应的数据集
深度学习识别方法	能获取更深层次的图像特征,图像特征表达更为丰富 无人为以及外界环境因素的干扰,图像特征提取更加准确 图像识别正确率较高	计算量较大,模型训练时间较长 硬件环境要求较高,耗资较大 模型训练过程中容易出现梯度消散和崩溃问题

对比精度较低,致使系统的整体识别效率无法达到预期要求。随着深度学习技术的发展,科研工作者将人脸识别与深度学习技术相结合,提出了一系列的人脸识别算法,较传统的人脸识别算法具有更高的识别精度。例如:

栗科峰等人<sup>[40]</sup>提出一种融合深度学习与最大间距准则的人脸识别算法,此算法有效改善了人脸识别技术遇到的突出问题,比传统的分类算法具有更高的准确度。但是,算法的不足在于出现重度遮挡、大范围特征图像变化以及图像分辨率降低时算法识别率较低。

Wu 等人<sup>[41]</sup>提出了一种优化的稀疏深度学习网络(SDLN)模型用于人脸识别(FR),利用SDLN学习基于局部结构化稀疏表示的人脸特征编码。在传统的稀疏FR方法中,训练样本的不同姿态和表达方式对识别结果有很大的影响。此算法对提出的SDLN进行了联合微调,以优化FR任务,其识别正确率在ORL和FERET数据库上分别达到98.5%和96.9%。但是,当不同类型的训练样本不足时,提取的图像特征欠缺,识别率降低。马晓等人<sup>[42]</sup>提出一种基于深度学习特征的稀疏表示的人脸识别方法。此方法首先利用深度CNN提取人脸特征,然后通过稀疏表示对所得人脸特征进行分类和识别。实验证明,此方法具有较好的识别准确度,较传统方法能够提升6%~60%,可以有效地应对光照、姿态、表情、遮挡等类内变化带来的干扰,且在小样本问题中具有较大的优势。但是,算法在实验过程出现图像线性特征表达不理想,通过改进网络结构和损失函数,使网络所得特征更加满足线性子空间。

Shailaja 等人<sup>[43]</sup>提出一种基于深度学习的线性判别回归分类(LDRC)算法,提高人脸识别(FR)系统的精度。在YALE和ORL数据库上的实验结果表明,该系统比早期的LRC算法具有更好的性能。在YALE数据库上其识别准确率达到92.8%,在ORL数据库上其识别准确率提升了20.3%。由于算法复杂度较高,系统运行耗时较长,算法结构有待进一步优化。

孔英会等人<sup>[44]</sup>提出一种基于CNN和CUDA加速的实时视频人脸识别算法,首先设计了一个CNN和Softmax的深度学习模型,然后结合CUDA加速算法实现人脸识别功能。结果证明,此方法可满足识别准确率和实时性要求,静态图像识别率在ORL和Yale数据库上分别达到96.5%和97.8%。在多姿态和环境变化较为剧烈的视频人脸识别中仍取得了较好的识别效果,但是,在对视频人脸识别过程中,未引入多级判决时,对于较为不确定的人脸强行输出识别结果,会导致识别错误。

李倩玉等人<sup>[45]</sup>提出一种改进的深层网络自动提取特征的人脸识别算法。此算法优点在于卷积核是通过单独的无监督学习获得,利用Softmax回归模型对提取的特征进行分类。结果证明,当编码器的级数为2时,算法在Yale、Yale-B和CMU-PIE数据库上的识别率分

别为97.8%、95.00%和96.17%。但是,当编码器级数增加时,识别率降低。

Fu 等人<sup>[46]</sup>提出了一种新型CNN体系结构——引导卷积神经网络(Guided-CNN),解决了在不同分辨率下人脸图像匹配度低下问题,在不同的测试方法中,均达到了较好的正确率,最高匹配正确率达到97.4%。

在基于深度学习的人脸识别发展过程中,有许多代表性的算法被提出,这些算法的性能在低分辨率的人脸图像上仍取得了较好的成绩。其具代表性的人脸识别算法在LFW与YTF数据库的正确识别率如表7所示。

表7 人脸识别算法在LFW与YTF数据库上的识别率

方法	%	
	LFW	YTF
Deep Face	97.35	91.40
FaceNet <sup>[47]</sup>	99.63	95.10
DeepID	97.47	—
DeepID2+ <sup>[48]</sup>	99.47	93.20
DeepID3 <sup>[49]</sup>	99.53	—
SphereFace <sup>[50]</sup>	99.42	95.00
LFRS <sup>[51]</sup>	97.73	92.24
DeepFR <sup>[52]</sup>	98.95	97.30
Center Face	99.28	94.90
Baidu <sup>[53]</sup>	99.77	—

随着CNN的兴起,人脸识别技术已成为深度学习重要研究领域之一,在此期间,Taigman 等人<sup>[54]</sup>提出的DeepFace和汤晓鸥团队提出的DeepID是人脸识别领域相对成功的模型。特别是汤晓鸥团队提出的DeepID模型,模型对CNN的内部结构进行了详细的分析,是CNN模型的一次本质性探索,具有重大科研意义,极大促进了人脸识别技术的发展。虽然,近几年人脸识别技术在各方面取得了不错的效果,但是其不足之处仍然相当明显,例如,如何正确识别双胞胎、怎样判断不同年龄段人脸的变化、怎样消除各种外部环境干扰(光照、遮盖物等)的影响、如何预防人脸被盗用并且通过人脸识别系统、如何缩短人脸识别模型的训练时间等。总之,解决上述问题,是当下人脸识别技术的进一步研究方向。

## 4.2 医学图像识别

医学图像具有种类繁多、分辨率低下、严重依赖成像设备和成像环境等缺点,这些缺点在一定程度上限制了医生对患者症状做出有效诊断。利用图像处理技术并结合深度学习实现对人体患病部位的诊断是当下最前沿的医学诊断方法之一。截止目前,仍有许多病变部位无法利用此技术对症状做出有效的诊断,主要原因是缺乏大量与之相关的数据集。现阶段利用此技术对病变部位进行诊断主要集中在脑部、胸部、腹部、眼部等方面,研究者在这些数据量较大的医学症状上建立了相应的数据集,例如:阿尔兹海默症、乳腺癌、肺癌、子宫癌、糖尿病视网膜病变等,并且在这些研究领域取得了一定



的研究成果。例如:

刘吉等人<sup>[55]</sup>使用深度学习方法对宫颈检查图像进行识别分类。首先对图像进行增大信噪比,并在图像输入模型之前进行随机变换以增大训练集,然后建立CNN模型,最后在测试集上图像识别准确率达到89.1%。不足之处在于原始数据较少,算法准确率较低。吕鸿蒙等人<sup>[56]</sup>提出了一种基于增强AlexNet的阿尔兹海默病的早期诊断,增加神经网络结构的层数以及优化各层参数,使之达到良好的诊断效果。实验结果证明,增强的AlexNet网络在诊断AD上更有优势,它的灵敏度达到100%,但它的特异度小于原始模型,说明增强的AlexNet在误诊率方面要大于原有模型。

李书通等人<sup>[57]</sup>提出一种非监督学习方法——3D-PCANet的阿尔兹海默病辅助诊断,通过非监督的方式学习数据之间的联系和特征,在数据集上达到了良好的效果,其平均分类正确率达到82.21%。但不足的是,算法使用的数据集是未经过数据筛选和数据排除,致使识别精度相对较低。

Mohamed等人<sup>[58]</sup>训练了一个CNN模型,利用所在机构收集的乳腺图像为数据集,对乳腺密度进行分类,确定乳腺癌发生的风险,实验表明,该方法的曲线下面积(Area Under Curve, AUC)达到98.8%。但对于两个难以区分的乳腺密度类别分类准确度较高时,为了准确判断乳腺癌发生的风险,这两种类别通常由放射科医生进行评估。

何雪英等人<sup>[59]</sup>提出一种改进的深度卷积神经模型,利用数据增强和迁移学习方法,实现对乳腺癌病理图像的自动分类,实验结果表明,该方法的识别率可达到91%。但数据增强之后,模型的训练时间增加。

Xu等人<sup>[60]</sup>提出一种堆叠式稀疏自编码器(SSAE),作为深度学习策略的一个实例,用于对乳腺癌的高分辨率组织病理图像进行有效的核检测,其准确率高达88.84%。但仅限于高分辨率组织病理图像识别。当病理图像像素较低时,尝试使用图像处理技术,丰富图像内容,达到对低分辨率图像的准确识别。

Coudray等人<sup>[61]</sup>利用癌症基因图谱,获得肺细胞组织病理图像,并且利用深度学习技术训练一个CNN模型,该模型能准确区分病变细胞与正常细胞,期预测结果的AUC达到97.0%。但此方法用于常见的癌症诊断,对于不常见的癌症诊断,如:大细胞癌、小细胞肺癌等,由于可用的数据量不足,无法正确预测细胞是否癌变,并且此方法在常规任务和困难病例(例如,区分肺内转移和多个同步原发性肺癌)中发挥作用较小,目前无法提供此方面的相关治疗决策。

李琮等人<sup>[62]</sup>提出一种基于卷积神经网络和分类器的视网膜病理图像自动分类系统,该方法首先对视网膜图像进行预处理,其次,在AlexNet的基础上,在每一个

卷积层和全连接层前引入一个批归一化层,得到一个更加复杂的深度卷积神经网络——BNnet。经测试,该网络在数据集上表现出较好的效果,其分类准确率达到93%。但在该实验中,图像预处理、特征提取和图像分类3个阶段尚未形成一个完整连续的系统,尝试改进网络结构进一步提高分类准确率。

翁铭等人<sup>[63]</sup>对基于深度学习的DR筛查智能诊断系统进行评价,该方法收集了2017年1至6月糖尿病患者的病理数据,总计186例372眼,其次比较专家诊断与基于深度学习的智能系统诊断,其结果显示,专家诊断与智能系统诊断具有高度的一致性,但是,智能系统诊断的灵敏度和特异性为82%和91%,明显优于人类专家。但系统依然存在不足之处,比如,研究中样本量较小、未采用以眼底血管造影为诊断参考标准等。同时,仅对DR分期程度研究,并未对视力有损害的DME进行筛查研究,实验结果有待进一步完善。

Ghesu等人<sup>[64]</sup>提出了一种基于边缘空间深度学习(MSDL)的图像目标检测与分割方法,利用层次边缘空间中有效对象参数化的优点,结合深度学习,网络自动设计新的特征框架。在对象定位的基础上,提出了一种基于深度学习的主动判别模型来估计非刚性物体边界。实验结果表明,与现有技术相比,新框架的准确率提高了45.2%。不足之处在于新架构中的大多数稀疏模式都达到了90%~95%的稀疏级别,因此在扫描过程中,索引的数据体量仅占总数据体量的一小部分,最大限度地减少了大容量数据流中的内存占用,虽然准确率提高,但是图像的深度特征并没有完全提取,这与CNN架构不同,CNN架构根据卷积核的大小对每个图像层进行多次采样。因此,将CNN整合到开发的混合框架中,是下一步的工作重点。

Wurfl等人<sup>[65]</sup>提出一种基于深度学习框架的三维断层重建方法。不同于大多数用于重建的深度学习方法,此方法允许在体积和投影域联合优化校正。实验结果表明,在感兴趣区域,较其他方法断层重建,其峰值信噪比增加了23%。但是,在断层重建过程中,许多伪影通常在投影域中出现。一方面是由光束硬化、散射和金属伪影等物理效应产生的伪影,另一方面是由数据丢失产生的伪影,如截断校正或有限角度重建等。如何改善或消除伪影现象,是下一阶段的研究重点。

Gao等人<sup>[66]</sup>提出一种能自动区分并识别白内障严重程度的系统,此系统构建了一个CNN和SVM相结合的学习模型,实验结果表明,该模型较传统的识别方法准确率提高了5.6%。但是,此系统识别白内障的严重程度严重依赖于人为对数据的标注是否完全正确,否则将会降低系统识别的正确率。对严重分割错误和人为错误标记进行分析可以进一步完善系统,提高识别精度。

Yan等人<sup>[67]</sup>设计了一个多阶段深度学习模型,用于

身体部位病变识别,在训练阶段,通过CNN来提取最具差异性的特征及切片中的局部信息;在增强阶段,经过预训练的CNN进一步增强图像的局部信息,实验结果证明,算法的准确率高达92.23%。但是,对三维图像检测时,其准确率无法达到预期要求,考虑使用三维卷积滤波器,对模型进行改进。

Sun等人<sup>[68]</sup>提了一种融合卷积神经网络(CNN)和支持向量机(SVM)两种高级分类器协同进行功能磁共振成像(fMRI)识别的混合模型,这两种分类器在图像识别领域均取得了良好的效果。在HaxBy的2001 FMRI数据集上进行实验,该融合方法获得了99.5%的识别准确率。

虽然,新型混合模型的识别率足够高,但是模型的性能有待提高,例如:可以对输入层的大小、第二层到第四层的特征层映射数以及模型中使用的内核函数进行改进。此外,进一步研究图像预处理算法,对提高分类精度至关重要。

Yu等人<sup>[69]</sup>提出一种将深度卷积神经网络(CNN)、Fisher向量(FV)和支持向量机(SVM)相结合的混合模型,用于皮肤图像评价。在ISBI 2016皮肤病挑战数据集上并取得了不俗的成果,较其他的识别方法,此方法的识别率提高3%。由于混合模型比一般的CNN具有更好的尺度不变性,因此,更适合于整体图像的表达,对于残缺的图像,模型提取图像特征较差,试着结合多尺度图像信息,提高残缺图像的识别精度。

图像识别技术在医学上主要应用在对患者病变部位的诊断,帮助医生快速对患者病变部位做出准确的诊断。随着CT、核磁共振、X射线等大型医疗影像设备的出现,图像识别技术成功应用在医学诊断上,但是,由于医疗行业的局限性,科研工作者们无法使用大量的医学图像作为原数据,因此,建立大规模的医学图像数据集、优化算法模型是医学图像识别的主要研究方向。

### 4.3 遥感图像分类

随着遥感技术的发展,遥感图像数据日益增加,其中,遥感图像中蕴含着大量有价值的信息,这些信息广泛应用于土地、水利、环境污染以及目标检测等方面。目前,遥感图像主要分为高光谱遥感图像和高分辨率遥感图像。由于传统的分类方法的局限性,例如,贝叶斯模型和K-均值等方法,无法对遥感图像进行准确的分类,随着深度学习的兴起,将深度学习技术应用到遥感图像的分类之中,与传统分类方法相比,取得了很好的分类效果。

王鑫等人<sup>[70]</sup>提出一种改进的基于深度学习的高分辨率遥感图像分类算法,首先,设计一个七层的CNN,将第五层池化输出,采用主成分分析进行降维,融合得到的高层特征,最后,设计一种逻辑分类器对图像进行分类。经测试,与传统基于深度学习的遥感图像分类方法相比,其准确率提高了8.66%。由于本文算法只基于

深度学习理论,算法效率有待提高,这也是未来进一步要改进和研究的方向。之后,王鑫等人<sup>[71]</sup>为了解决传统遥感图像分类方法特征提取过程复杂、特征表现力不强等问题,提出了一种基于卷积神经网络和多核学习的高分辨率遥感图像分类方法,设计了一种基于CNN-多核学习-SVM模式的遥感图像分类器,对遥感图像进行分类,实验结果表明,该算法在分类正确率、误分类率和Kappa系数等系能指标均有所提升,在测试集上分标达到了96.43%、3.57%和96.25%。

Zhang等人<sup>[72]</sup>提出一种基于对象的深度学习生态环境要素分类方法。首先,训练一个深度卷积神经网络(DCNN)来区分不同的生态环境元素。然后,通过训练的DCNN提取这些子图像的深度特征。之后,使用SOFTMax分类器来预测所有子图像的概率。在测试集上,该方法的分类精度达到了98.44%。此外,不规则形状区域的分类精度也达到96.77%。算法不足之处在于那些高度相似的生态环境难以区分,从图像层面上观测,一些元素非常相似,如裸地和湿地等。

Huang等人<sup>[73]</sup>提出一种基于深度反卷积神经网络(DEPONVNET)的有监督建筑物提取方法。首先创建一个高质量的建筑物数据集,然后在一个公共的大型建筑数据集上预演一个深度反卷积网络,此外,精细调整模型的输出。在建筑数据集上的大量实验证明了所提出的方法具备有效性。但是,在预训练深度反卷积网络模型时,耗时较长。

Zhou等人<sup>[74]</sup>提出一种基于卷积神经网络(CNN)的Landsat-8多光谱遥感图像花生种植面积提取方法。实验结果表明,所设计的CNN结构具有良好的性能,总体精度为96.42%,Kappa效率为0.944。实验结果表明,利用CNN进行多光谱图像作物识别的探索,在农业遥感分类中具有优异的性能和巨大的潜力。但是,作者只考虑了作物的光谱信息,应该更多地关注作物在时间阶段的信息等,除此之外,试着提取光谱空间信息的相关性,以提高分类精度。

Lin等人<sup>[75]</sup>提出一种基于HOG和深度学习特征的遥感图像飞机识别方法。训练了两种分类器,一种是基于HOG特征的SVM分类器,另一种是基于深卷积神经网络VGGNet分类器。首先利用SVM分类器对图像中的飞机进行粗略识别,然后利用深度学习分类器对误识别的目标进行排除。此方法比传统方法具有更好的泛化能力。其准确率高达98.1%。

Zou等人<sup>[76]</sup>提出一种基于特征重构的深度学习方法对遥感图像进行分类,使用深层信念网络(DBN)在整个特征集上最小化重构误差来实现特征抽取,所提出的方法选择更具可重构性的特征作为判别特征。实验结果证明,算法的最高识别精度高达93.5%。算法不足在于对已混淆的场景观测精度较低,在训练过程中,训练数据超过50%时,算法的性能有所降低。



表8 遥感图像算法对比分析

遥感图像	CNN	DBN	SAE
高光谱遥感图像	无法直接分类,需要降维处理	较好地提出图像的深层特征,提高分类精度	可以降低维处理,结合极限学习机的方法解决数据复杂和样本数目有限问题
高分辨率遥感图像	更好地获取影像的纹理,空间特征等信息	更加准确地提取出遥感影像的空间分布规律	解决 SAR 影像特征表示较少和斑点噪声存在的问题

Yang 等人<sup>[77]</sup>提出一种多视图 CCANETs(MCCANETs) 遥感图像识别。MCCANETs 模型通过 MCCA 方法学习层叠的多透视滤波器组,建立了一个深层卷积结构。在输出阶段,分别采用二值化和分块直方图作为非线性处理的特征池。在 RSSCN7 RS 数据库上进行了大量的实验,结果表明,MCCANETs 优于双视图 CCANET,其识别率提高 2.12%。但当训练次数达到 2 000 次时,算法识别率有所降低,其次随着卷积层层数增加时,MCCANETs 的识别率的增长率较 CCANETs 有所降低。

Deng 等人<sup>[78]</sup>提出一种改进的 CNN,解决了检测密集堆积的物体识别率低下的问题,方法基于最新的残余网络(ResNets)所提出,由两个子网络组成:一个对象建议网络和一个对象检测网络。为了检测密集堆积的物体,将多尺度层的输出组合在一起,提高特征图的分辨率,在测试集上取得了较好的效果。由于算法需要提取更高层、更细微的图像信息,导致网络结构层数更多,耗时较长。试着结合多 GPU 配置,进一步减少大规模遥感图像的计算时间。其次,算法对大小可变的对象检测效果不理想。

Tong 等人<sup>[79]</sup>提出一种改进的 DBN 高光谱遥感图像分类方法,方法主要改进了 DBN 的训练流程,在 Salinas 数据集对该方法进行验证,实验结果表明,与 BP 网络、SVM、标准训练过程的 DBN 相比,该模型具有较好的分类效果,当训练样本占数据集的 40% 时,模型的精确度可达到 96.72%。但是,设计的 DBN 存在一个缺点,即很容易混淆结构相似的对象。例如,数据集中对不同等级葡萄对象的识别效果不理想。

Tanase 等人<sup>[80]</sup>提出一种自定义卷积 DBN 模型,用于全极化 SAR 数据特征提取,并且逐层提取更高层次的遥感图像特征。利用 CNN 扩展到全尺寸的 PolSAR 图像,能够有效地提取出 SAR 影像特征。最后在对德国的 Kaufbeuren 机场的全极化多视 F-SAR 影像的斑块提取中,算法得到有效的验证。

Lv 等人<sup>[81]</sup>提出一种基于叠加自编码(SAE)的遥感图像分类算法,首先,对样本数据采用样本分割和 SAE 技术,从而在基本分类器中产生高度的多样性,其次,选择 ELM(Extreme Learning Machine)神经网络作为基本的分类器,用来提高算法的计算速度,最后,采用 Q 统计量确定最终集合的分类器。实验将该算法与 Bagging、Adaboost、Random Forest 等算法进行比较分析,结

果表明,该算法不仅在低分辨率、中分辨率、高分辨率和高光谱遥感图像中具有较高的分类精度,而且对 UCI 数据具有较强的稳定性和泛化性。但是,无法确定 Q 统计量的基本分类器的数目对算法性能是否具有影响,目前仅采用人工选择,理论依据欠缺。

然而对高分辨率遥感图像与高光谱遥感图像的识别分类最常用神经模型是 CNN、DBN 以及 SAE,两种遥感图像在三种算法在的对比分析如表 8 所示。

目前,遥感图像分类识别主要依赖于人工识别,遥感图像的数据主要来源于卫星影像,由于遥感图像的数据过于庞大,容易产生信息冗余,并且图像分辨率较低,容易导致不同信息相融合现象。由于遥感图像的本质特征,导致对遥感图像的分类较为困难,因此,如何准确分离出具有价值的信息,是遥感图像分类的重要任务。利用深度学习技术理论上是可以实现对遥感图像信息的分类,但是,由于技术上的不足,距离实际应用依然存在一定的差距。针对遥感图像的特殊性建立合适的深度学习模型,使用较好的优化算法,其识别率几乎接近人工识别,这是其遥感图像分类识别的发展趋势。

#### 4.4 其他领域识别

##### 4.4.1 ImageNet 分类识别

ImageNet ILSVRC 是图像识别领域的一个主要体现。在 2012 年之后,Hinton<sup>[82]</sup> 以及其学生采用更深的 CNN 模型在 ImageNet 挑战赛上取得了最好的结果。主要对算法做了进一步的优化,并且在训练 CNN 模型时使用了权重衰减的方法,防止 CNN 模型在训练过程中出现过拟合现象。另一方面由于大型计算机的产生,计算能力大幅度提高,GPU 等硬件设施的不断完善,使网络能够更好地拟合训练数据,促使对图像识别的领域研究更进一步。在历年的 ImageNet ILSVRC 挑战赛中,以深度学习技术为基础的算法模型在识别率上领先于其他算法模型,表 9 为历年在 ImageNet ILSVRC 挑战赛中,图像分类算法的冠军模型。

表9 历年 ImageNet ILSVRC 图像分类算法模型冠军

年份	结构模型	模型深度/层	Top5 错误率/%
2012	AlexNet	8	15.315
2013	ZFNet	8	11.197
2014	GoogLeNet	22	6.656
2015	ResNet	152	3.570
2016	GoogLeNet WideResNet	64	2.991
2017	SENet	152	2.251

由表9可知,近几年的冠军模型均是以CNN模型为基础,深度优化发展而来,正是由于CNN独特的结构,使CNN在图像识别方面优于其他神经网络模型。从历年冠军模型的Top5错误率上可知,改进的CNN模型在识别准确率上已经超越了人类。

#### 4.4.2 交通识别

在交通领域,图像识别技术广泛应用于车牌识别、车道偏移预警、交通标志识别等方面。该技术在交通控制、智能停车、收费管理和安全等智能城市建设中发挥着重要作用,这也是图像识别技术应用最早、技术最成熟的领域之一<sup>[83]</sup>。近几年,研究者将深度学习技术运用到交通领域,并且在此方面取得了一定的科研成果,例如,Chang等人<sup>[84]</sup>提出一种基于深度卷积神经网络(CNN)的交通标志检测算法,该算法具有实时检测速度和99%以上的检测精度。虽然该算法识别率较高,但是算法模型训练较慢,时间较长。针对此现象,Zou等人<sup>[85]</sup>使用一种更快的R-CNN检测交通标志算法,虽然此算法没有达到更理想的效果,但是训练速度更快,为在交通标志检测方面的探索提供了一种新的方法。

#### 4.4.3 字符识别

字符识别是图像识别技术应用最早的领域之一,在支票、邮政信封、电子签名等方面有着广泛用途,最早的字符识别基于模式识别技术,需要人为对字符进行预处理,这样识别效率较低。随着深度学习技术的兴起,科研工作者开始使用深度学习技术识别字符,Ciresan等人<sup>[86]</sup>使用多个CNNs识别手写字符,在MNIST数据集上取得了最好的成绩,使错误率降到了0.27%。之后,Wu等人<sup>[87]</sup>提出了一种基于松弛卷积神经网络(R-CNN)和交替训练松弛卷积神经网络(ATR-CNN)的手写体识别方法,在ICDAR13中国手写字符识别竞赛中取得冠军,错误率为3.94%,进一步缩小了机器和人类观察者之间的差距。虽然近几年深度学习技术在字符识别方面取得了巨大成功,但是,不足之处还是较为明显,如何正确识别相似字符、稀有字符、随意性字符等,是未来字符识别应解决的问题。

## 5 结束语

通过前文的回顾与总结,可以知道科研工作者在深度学习与图像识别的交叉领域已取得了诸多成果,然而当前机器学习在图像识别中的应用研究还不尽完善,有值得改进的空间,现通过本次综述,列举未来可能的研究重点与方向,期望推动本领域研究的发展。

#### (1) 基于小样本数据集的迁移学习识别方法研究

目前,图像识别算法和模型的验证是以大量样本数据为依托,建立起相应的数据集,如:字符数据集、人脸数据集、车辆数据集等。但是,在日常生活中,对其他事物的识别是缺乏大样本数据集,如:火灾数据集等。虽

然不能把每一种类型的数据制作成一个大型数据集,但可以采集与之相关的少量样本作为识别数据集。因此,有效使用迁移学习技术识别小样本数据的方法,是未来科的研究方向之一。

#### (2) 加强非监督与半监督学习识别的研究

监督学习是图像识别领域应用最广泛的深度学习方法,但是,监督学习的训练需要带标签的数据,由于标签需要人为添加的,会消耗大量的人力、物力和财力。而非监督学习与半监督学习弥补了监督学习的不足。目前虽然非监督学习与半监督学习在图像识别方面取得了一定的进展,但是,和监督学习相比,非监督学习与半监督学习在图像识别精确度方面依然差距较大。

#### (3) 加强深度学习在视频图像识别中的研究

相对于静态图像识别,深度学习在视频图像方面识别应用欠缺。一方面由于缺少大量的视频图像数据集,另一方面训练视频图像需要更深层次的模型,模型结构更为复杂,计算量更大,训练时间将会延长。如何改善这些问题,是视频图像识别研究的关键之一。

#### (4) 强化深度学习模型的理论研究

深度学习模型通过大量数据训练之后,能极好地反映出图像的特征。但是,深度学习模型的理论性研究仍然有许多未完成的工作,模型的理论基础有待进一步坚实。例如,模型何时收敛?怎样更好地找到极小值点?图像信息进入模型下一层时,其损失的信息有哪些?

## 参考文献:

- [1] 郭丽丽,丁世飞.深度学习研究进展[J].计算机科学,2015,42(5):28-33.
- [2] Hinton G E.A practical guide to training restricted boltzmann machines[J].Momentum,2012,9(1):599-619.
- [3] Hinton G E,Salakhutdinov R R.Reducing the dimensionality of data with neural networks[J].Science,2006,313(5786):504-507.
- [4] Hinton G E,Osindero S,Teh Y W.A fast learning algorithm for deep belief nets[J].Neural Computation,2006,7(18):1527-1554.
- [5] Babri H A,Tong Y.Deep feedforward networks: application to pattern recognition[C]//Proceedings of International Conference on Neural Networks(ICNN'96),1996:1422-1426.
- [6] Lecun Y,Bottou L,Bengio Y,et al.Gradient-based learning applied to document recognition[J].Proceedings of IEEE,1998,86(11):2278-2324.
- [7] Shao L,Wu D,Li X.Learning deep and wide:a spectral method for learning deep networks[J].IEEE Transactions on Neural Networks and Learning Systems,2014,25(12):2303-2308.
- [8] 胡清华,张道强,张长水.复杂环境下的机器学习研究专刊前言[J].软件学报,2017(11):2811-2813.



- [9] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [10] Schapire R E. Theoretical views of boosting and applications[C]//Proceedings of 10th International Conference on Algorithmic Learning Theory (ALT 99). Berlin: Springer, 1999: 13-25.
- [11] Krishnapuram B, Carin L, Figueiredo M A T, et al. Sparse multinomial logistic regression: fast algorithms and generalization bounds[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(6): 957-968.
- [12] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504.
- [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.
- [14] 刘方园, 王水花, 张煜东. 深度置信网络模型及应用研究综述[J]. 计算机工程与应用, 2018, 54(1): 11-18.
- [15] Zhong P, Gong Z, Li S, et al. Learning to diversify deep belief networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(6): 3516-3530.
- [16] Huang G, Liu Z, Maaten L V D, et al. Densely connected convolutional networks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261-2269.
- [17] Zhang K, Guo Y, Wang X, et al. Multiple feature reweight DenseNet for image classification[J]. IEEE Access, 2019, 7: 9872-9880.
- [18] 周俊宇, 赵艳明. 卷积神经网络在图像分类和目标检测应用综述[J]. 计算机工程与应用, 2017, 53(13): 34-41.
- [19] Lipton Z C. A critical review of recurrent neural networks for sequence learning[EB/OL]. (2015)[2019]. <https://arxiv.org/abs/1506.00019v1>.
- [20] 朱小燕, 王昱, 徐伟. 基于循环神经网络的语音识别模型[J]. 计算机学报, 2001(2): 213-218.
- [21] 张剑, 屈丹, 李真. 基于词向量特征的循环神经网络语言模型[J]. 模式识别与人工智能, 2015(4): 299-305.
- [22] Mirza A H, Cosan S. Computer network intrusion detection using sequential LSTM Neural Networks autoencoders[C]//Proceedings of 26th Signal Processing and Communications Applications Conference, 2018: 1-4.
- [23] Cho K, van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL]. (2014)[2019]. <https://arxiv.org/abs/1406.1078>.
- [24] 郝志峰, 黄浩, 蔡瑞初, 等. 基于多特征融合与双向RNN的细粒度意见分析[J]. 计算机工程, 2018(7): 199-204.
- [25] Mou L, Ghamisi P, Zhu X. Deep recurrent neural networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3639-3655.
- [26] 陈欣于, 俊洋, 赵媛媛. 基于CNN和B-LSTM的文本处理模型研究[J]. 轻工学报, 2018, 33(5): 103-108.
- [27] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014: 2672-2680.
- [28] Zhu L, Chen Y, Ghamisi P, et al. Generative adversarial networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(9): 5046-5063.
- [29] 商显震, 韩萌, 孙毓忠, 等. 融合生成对抗网络和朴素贝叶斯皮肤病诊断方法[J]. 计算机科学与探索, 2019, 13(6): 1005-1015.
- [30] 唐贤伦, 杜一铭, 刘雨微, 等. 基于条件深度卷积生成对抗网络的图像识别方法[J]. 自动化学报, 2018, 44(5): 855-864.
- [31] Mehdi Mirza S O. Conditional generative adversarial nets[EB/OL]. (2014)[2019]. <https://arxiv.org/abs/1411.1784>.
- [32] Denton E, Chintala S, Szlam A, et al. Deep generative image models using a laplacian pyramid of adversarial networks[J]. Computer Science, 2015.
- [33] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[EB/OL]. (2017)[2019]. <https://arxiv.org/abs/1701.07875>.
- [34] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs[EB/OL]. (2017)[2019]. <https://arxiv.org/abs/1704.00028>.
- [35] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. Computer Science, 2015.
- [36] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks[C]//Proceedings of IEEE Conference on Computer Vision, Venice, Italy, 2017: 2813-2821.
- [37] Berthelot D, Schumm T, Metz L. BEGAN: boundary equilibrium generative adversarial networks[EB/OL]. (2017)[2019]. <https://arxiv.org/abs/1703.10717>.
- [38] Sabour S, Frosst N E, Hinton G. Dynamic routing between capsules[EB/OL]. (2017)[2019]. <https://arxiv.org/abs/1710.09829>.
- [39] Deng F, Pu S, Chen X, et al. Hyperspectral image classification with capsule network using limited training samples[J]. Sensors, 2018, 18(9): 22.
- [40] 栗科峰, 黄全振. 融合深度学习与最大间距准则的人脸识别方法[J]. 计算机工程与应用, 2018, 54(5): 206-210.
- [41] Wu R, Kamata S. A jointly local structured sparse deep learning network for face recognition[C]//Proceedings of IEEE International Conference on Image Processing (ICIP). Phoenix: IEEE, 2016: 3026-3030.
- [42] 马晓, 张番栋, 封举富. 基于深度学习特征的稀疏表示的人脸识别方法[J]. 智能系统学报, 2016(3): 279-286.
- [43] Shailaja K, Anuradha B. Effective face recognition using

- deep learning based linear discriminant classification[C]//Proceedings of IEEE International Conference on Computational Intelligence and Computing Research (ICIC). Chennai:IEEE,2016:1-6.
- [44] 孔英会,王之涵,车麟麟.基于卷积神经网络(CNN)和CUDA加速的实时视频人脸识别[J].科学技术与工程,2016(35):96-100.
- [45] 李倩玉.基于改进深层网络的视频人脸识别研究[D].合肥:合肥工业大学,2016.
- [46] Fu T,Chiu W,Wang F.Learning guided convolutional neural networks for cross-resolution face recognition[C]//Proceedings of 27th International Workshop on Machine Learning for Signal Processing (MLSP).Tokyo:IEEE,2017:1-5.
- [47] Schroff F,Kalenichenko D,Philbin J.FaceNet:a unified embedding for face recognition and clustering[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR).Boston:IEEE,2015:815-823.
- [48] Sun Y,Wang X,Tang X.Deeply learned face representations are sparse,selective,and robust[J].Computer Science,2014.
- [49] Sun Y,Liang D,Wang X,et al.DeepID3:face recognition with very deep neural networks[J].Computer Science,2015.
- [50] Liu W,Wen Y,Yu Z,et al.SphereFace:deep hypersphere embedding for face recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR).Honolulu:IEEE,2017:6738-6746.
- [51] Yi D,Lei Z,Liao S,et al.Learning face representation from scratch[J].Computer Science,2014.
- [52] Omkar M,Parkhi A V A A.Deep face recognition[C]//British Machine Vision Conference(BMVC),2015:1-12.
- [53] Liu J,Deng Y,Bai T,et al.Targeting ultimate accuracy:face recognition via deep embedding[J].Computer Science,2015.
- [54] Taigman Y,Yang M,Ranzato M,et al.DeepFace:closing the gap to human-level performance in face verification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.Columbus:IEEE,2014:1701-1708.
- [55] 刘吉,孙仁诚,乔松林.深度学习在医学图像识别中的应用研究[J].青岛大学学报(自然科学版),2018(1):69-74.
- [56] 吕鸿蒙,赵地,迟学斌.基于增强 AlexNet 的深度学习阿尔茨海默病的早期诊断[J].计算机科学,2017,44(S1):50-60.
- [57] 李书通,肖斌,李伟生,等.基于 3D-PCANet 的阿尔兹海默病辅助诊断[J].计算机科学,2018,45(S1):140-142.
- [58] Mohamed A A,Berg W A,Peng H,et al.A deep learning method for classifying mammographic breast density categories[J].Medical Physics,2018,45(1):314-321.
- [59] 何雪英,韩忠义,魏本征.基于深度学习的乳腺癌病理图像自动分类[J].计算机工程与应用,2018,54(12):121-125.
- [60] Xu J,Xiang L,Liu Q,et al.Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images[J].IEEE Transactions on Medical Imaging,2016,35(1):119-130.
- [61] Coudray N,Moreira A L,Sakellaropoulos T,et al.Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning[J].bioRxiv,2017:197574.
- [62] 李琼,柏正尧,刘莹芳.糖尿病性视网膜图像的深度学习分类方法[J].中国图象图形学报,2018,23(10):1594-1603.
- [63] 翁铭,郑博,吴茂念,等.基于深度学习的 DR 筛查智能诊断系统的初步研究[J].国际眼科杂志,2018,18(3):568-571.
- [64] Ghesu F,Krubasik E,Georgescu B,et al.Marginal space deep learning:efficient architecture for volumetric image parsing[J].IEEE Transactions on Medical Imaging,2016,35(5):1217-1228.
- [65] Wurfl T,Hoffmann M,Christlein V,et al.Deep learning computed tomography:learning projection-domain weights from image domain in limited angle problems[J].IEEE Transactions on Medical Imaging,2018,37(6):1454-1463.
- [66] Gao X,Lin S,Wong Y.Automatic feature learning to grade nuclear cataracts based on deep learning[J].IEEE Transactions on Biomedical Engineering,2015,62(11):2693-2701.
- [67] Yan Z,Zhan Y,Peng Z,et al.Multi-instance deep learning:discover discriminative local anatomies for bodypart recognition[J].IEEE Transactions on Medical Imaging,2016,35(5):1332-1343.
- [68] Sun X,Park J,Kang K,et al.Novel hybrid CNN-SVM model for recognition of functional magnetic resonance images[C]//Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC).Banff:IEEE,2017:1001-1006.
- [69] Yu Z,Ni D,Chen S,et al.Hybrid dermoscopy image classification framework based on deep convolutional neural network and Fisher vector[C]//Proceedings of 14th International Symposium on Biomedical Imaging (ISBI 2017).Melbourne:IEEE,2017:301-304.
- [70] 王鑫,李可,徐明君,等.改进的基于深度学习的遥感图像分类方法[J].计算机应用,2019,32(2):382-387.
- [71] 王鑫,李可,宁晨,等.基于深度卷积神经网络和多核学习的遥感图像分类方法[J/OL].电子与信息学报.doi:10.11999/JEIT80628.
- [72] Zhang H,Yang B,Fang T,et al.Learning deep features for classification of typical ecological environmental elements in high-resolution remote sensing images[C]//Proceedings of 10th International Symposium on Computational Intelligence and Design (ISCID).Hangzhou:IEEE,2017:223-227.



- [73] Huang Z, Cheng G, Wang H, et al. Building extraction from multi-source remote sensing images via deep deconvolution neural networks[C]//Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing: IEEE, 2016: 1835-1838.
- [74] Zhou Z, Li S. Peanut planting area change monitoring from remote sensing images based on deep learning[C]//Proceedings of 4th International Conference on Systems and Informatics (ICSAI). Hangzhou: IEEE, 2017: 1358-1362.
- [75] Lin J, Li X, Pan H. Aircraft recognition in remote sensing images based on deep learning[C]//Proceedings of 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC). Nanjing: IEEE, 2018: 895-899.
- [76] Zou Q, Ni L, Zhang T, et al. Deep learning based feature selection for remote sensing scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(11): 2321-2325.
- [77] Yang X, Liu W, Tao D, et al. Multiview canonical correlation analysis networks for remote sensing image recognition[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1855-1859.
- [78] Deng Z, Lei L, Sun H, et al. An enhanced deep convolutional neural network for densely packed objects detection in remote sensing images[C]//Proceedings of International Workshop on Remote Sensing with Intelligent Processing (RSIP). Shanghai: IEEE, 2017: 1-4.
- [79] Tong G, Li Y, Cao L, et al. A DBN for hyperspectral remote sensing image classification[C]//Proceedings of 12th IEEE Conference on Industrial Electronics and Applications (ICIEA). Siem Reap: IEEE, 2018: 1757-1762.
- [80] Tanase R, Datcu M, Randucanu D. A convolutional deep belief network for polarimetric SAR data feature extraction[C]//The Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing: IEEE, 2016: 7545-7548.
- [81] Lv F, Han M, Qiu T. Remote sensing image classification based on ensemble extreme learning machine with stacked autoencoder[J]. IEEE Access, 2017, 5: 9021-9031.
- [82] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing System, 2012: 1097-1105.
- [83] 肖艳秋, 杜江恒, 闻萌莎, 等. 基于颜色特征和改进支持向量机算法的交通标志检测与识别[J]. 轻工学报, 2018, 33(3): 57-65.
- [84] Xiong C, Wang C, Ma W, et al. A traffic sign detection algorithm based on deep convolutional neural network[C]//Proceedings of IEEE International Conference on Signal and Image Processing (ICSIP). Beijing: IEEE, 2016: 676-679.
- [85] Zuo Z, Yu K, Zhou Q, et al. Traffic signs detection based on faster R-CNN[C]//Proceedings of IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW). Atlanta: IEEE, 2017: 286-288.
- [86] Ciresan C, Meier U, Gambardella M, et al. Convolutional neural network committees for handwritten character classification[C]//Proceedings of International Conference on Document Analysis and Recognition. Beijing: IEEE, 2011: 1135-1139.
- [87] Wu C, Fan W, He Y, et al. Handwritten character recognition by alternately trained relaxation convolutional neural network[C]//Proceedings of 14th International Conference on Frontiers in Handwriting Recognition. Heraklion: IEEE, 2014: 291-296.
- (上接第19页)
- [24] Yang B, Yih W T, He X, et al. Embedding entities and relations for learning and inference in knowledge bases[C]//Proceedings of the International Conference on Learning Representations, 2015: 233-240.
- [25] Neelakantan A, Roth B, McCallum A. Compositional vector space models for knowledge base completion[J]. Computer Science, 2015: 1-16.
- [26] Lin Y, Liu Z, Sun M. Modeling relation paths for representation learning of knowledge bases[J]. arXiv preprint arXiv: 1506.00379, 2015.
- [27] Das R, Neelakantan A, Belanger D, et al. Chains of reasoning over entities, relations, and text using recurrent neural networks[C]//Conference of the European Chapter of the Association for Computational Linguistics, 2017.
- [28] Shen Y, Huang P S, Chang M W, et al. Modeling large-scale structured relationships with shared memory for knowledge base completion[C]//Proceedings of the 2nd Workshop on Representation Learning for NLP. ACL, 2017: 57-68.
- [29] Graves A, Wayne G, Reynolds M, et al. Hybrid computing using a neural network with dynamic external memory[J]. Nature, 2016, 538(7626): 471-476.
- [30] Greff K, Srivastava R K, Koutník J, et al. LSTM: a search space odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28(10): 2222-2232.
- [31] Gers F A, Schmidhuber E. LSTM recurrent networks learn simple context-free and context-sensitive languages[J]. IEEE Transactions on Neural Networks, 2001, 12(6): 1333-1340.
- [32] Wang Z, Chen T, Ren J, et al. Deep reasoning with knowledge graph for social relationship understanding[J]. arXiv preprint arXiv: 1807.00504, 2018.
- [33] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61.
- [34] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//International Conference on Neural Information Processing Systems, 2015.