图像识别中的深度学习

王晓刚
香港中文大学

关键词：深度学习　物体识别　物体检测　视频分析

深度学习发展历史

深度学习是近十年来人工智能领域取得的重要突破。它在语音识别、自然语言处理、计算机视觉、图像与视频分析、多媒体等诸多领域的应用取得了巨大成功。现有的深度学习模型属于神经网络。神经网络的起源可追溯到20世纪40年代，曾经在八九十年代流行。神经网络试图通过模拟大脑认知的机理解决各种机器学习问题。1986年，鲁梅尔哈特(Rumelhart)、欣顿(Hinton)和威廉姆斯(Williams)在《自然》杂志发表了著名的反向传播算法用于训练神经网络[1]，该算法直到今天仍被广泛应用。

神经网络有大量参数，经常发生过拟合问题，虽然其识别结果在训练集上准确率很高，但在测试集上效果却很差。这是因为当时的训练数据集规模都较小，加之计算资源有限，即便是训练一个较小的网络也需要很长的时间。与其他模型相比，神经网络并未在识别准确率上体现出明显的优势。

因此更多的学者开始采用支持向量机、Boosting、最近邻等分类器。这些分类器可以用具有一个或两个隐含层的神经网络模拟，因此被称为浅层机器学习模型。在这种模型中，往往是针对不同的任务设计不同的系统，并采用不同的手工设计的特征。例物体识别采用尺度不变特征转换(Scale Invariant Feature Transform, SIFT)，人脸识别采用局部二值模式(Local Binary Patterns, LBP)，行人检测采用方向梯度直方图(Histogram of Oriented Gradient, HOG)特征。

2006年，欣顿提出了深度学习。之后深度学习在诸多领域取得了巨大成功，受到广泛关注。神经网络能够重新焕发青春的原因有几个方面：首先，大规模训练数据的出现在很大程度上缓解了训练过拟合的问题。例如，ImageNet[2]训练集拥有上百万个有标注的图像。其次，计算机硬件的飞速发展为其提供了强大的计算能力，一个GPU芯片可以集成上千个核。这使得训练大规模神经网络成为可能。第三，神经网络的模型设计和训练方法都取得了长足的进步。例如，为了改进神经网络的训练，学者提出了非监督和逐层的预训练，使得在利用反向传播算法对网络进行全局优化之前，网络参数能达到一个好的起始点，从而在训练完成时能达到一个较好的局部极小点。

深度学习在计算机视觉领域最具影响力的突破发生在2012年，欣顿的研究小组采用深度学习赢得了ImageNet图像分类比赛的冠军[3]。排名第2到第4位的小组采用的都是传统的计算机视觉方法、手工设计的特征，他们之间准确率的差别不超过1%。欣顿研究小组的准确率超出第二名10%以上，(见表1)。这个结果在计算机视觉领域产生了极大

表1　2012年ImageNet图像分类竞赛结果

排名	小组	Top5错误率	描述
1	多伦多大学	0.15315	深度学习
2	东京大学	0.26172	手工设计的特征
3	牛津大学	0.26979	
4	Zerox/INRIA	0.2758	

的震动，引发了深度学习的热潮。

计算机视觉领域另一个重要的挑战是人脸识别。有研究表明[5]，如果只把不包括头发在内的人脸的中心区域给人看，人眼在户外脸部检测数据库 (Labeled Faces in the Wild, LFW) 上的识别率是 97.53%。如果把整张图像，包括背景和头发给人看，人眼的识别率是 99.15%。经典的人脸识别算法 Eigenface [6] 在 LFW 测试集上只有 60% 的识别率。在非深度学习算法中，最高的识别率是 96.33% [7]。目前深度学习可以达到 99.47% 的识别率 [8]。

在欣顿的科研小组赢得 ImageNet 比赛冠军之后的 6 个月，谷歌和百度都发布了新的基于图像内容的搜索引擎。他们采用深度学习模型，应用在各自的数据上，发现图像搜索准确率得到了大幅度提高。百度在 2012 年成立了深度学习研究院，2014 年 5 月又在美国硅谷成立了新的深度学习实验室，聘请斯坦福大学著名教授吴恩达担任首席科学家。脸谱于 2013 年 12 月在纽约成立了新的人工智能实验室，聘请深度学习领域的著名学者、卷积网络的发明人雅恩·乐昆 (Yann LeCun) 作为首席科学家。2014 年 1 月，谷歌抛出四亿美金收购了深度学习的创业公司 DeepMind。鉴于深度学习在学术界和工业界的巨大影响力，2013 年，《麻省理工科技评论》(*MIT Technology Review*) 将其列为世界十大技术突破之首。

深度学习有何与众不同?

深度学习和其他机器学习方法相比有哪些关键的不同点，它为何能在许多领域取得成功?

特征学习

深度学习与传统模式识别方法的最大不同在于它所采用的特征是从大数据中自动学习得到，而非采用手工设计。好的特征可以提高模式识别系统的性能。过去几十年，在模式识别的各种应用中，手工设计的特征一直处于统治地位。手工设计主要依靠设计者的先验知识，很难利用大数据的优势。由于依赖手工调参数，因此特征的设计中所允许出现的参数数量十分有限。深度学习可以从大数据中自动学习特征的表示，可以包含成千上万的参数。

采用手工设计出有效的特征往往需要五到十年时间，而深度学习可以针对新的应用从训练数据中很快学习到新的有效的特征表示。

一个模式识别系统包括特征和分类器两部分。在传统方法中，特征和分类器的优化是分开的。而在神经网络的框架下，特征表示和分类器是联合优化的，可以最大程度地发挥二者联合协作的性能。

2012 年欣顿参加 ImageNet 比赛所采用的卷积网络模型 [9] 的特征表示包含了从上百万样本中学习得到的 6000 万个参数。从 ImageNet 上学习得到的特征表示具有非常强的泛化能力，可以成功应用到其他数据集和任务中，例如物体的检测、跟踪和检索等。在计算机视觉领域另外一个著名的竞赛是 PSACAL VOC。但是它的训练集规模较小，不适合训练深度学习模型。有学者将 ImageNet 上学习得到的特征表示用于 PSACAL VOC 上的物体检测，检测率提高了 20%[10]。

既然特征学习如此重要，那么，什么是好的特征呢？一幅图像中，各种复杂的因素往往以非线性的方式结合在一起。例如人脸图像中就包含了身份、姿态、年龄、表情、光线等各种信息。深度学习的关键就是通过多层非线性映射将这些因素成功分开，例如在深度模型的最后一个隐含层，不同神经元代表了不同因素。如果将这个隐含层当作特征表示，人脸识别、姿态估计、表情识别、年龄估计就会变得非常简单，因为各个因素之间变成了简单的线性关系，不再彼此干扰。

深层结构的优势

深度学习模型的“深”字意味着神经网络的结构深，由很多层组成。而支持向量机和 Boosting 等其他常用的机器学习模型都是浅层结构。三层神经网络模型（包括输入层、输出层和一个隐含层）可以近似任何分类函数。既然如此，为什么需要深层模型呢?

研究表明，针对特定的任务，如果模型的深度不够，其所需要

的计算单元会呈指数增加。这意味着虽然浅层模型可以表达相同的分类函数，但其需要的参数和训练样本要多得多。浅层模型提供的是局部表达。它将高维图像空间分成若干个局部区域，每个局部区域至少存储一个从训练数据中获得的模板，如图 1(a) 所示。

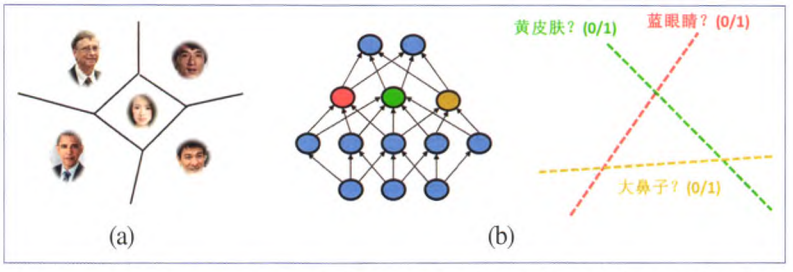

图1 (a)浅层模型得到的是局部表达；(b)深度模型提供分布式的特征表示

浅层模型将一个测试样本和这些模板逐一匹配，根据匹配的结果预测其类别。例如，在支持向量机模型中，模板是支持向量；在最近邻分类器中，模板是所有的训练样本。随着分类问题复杂度的增加，需要将图像空间划分成越来越多的局部区域，因而需要越来越多的参数和训练样本。尽管目前许多深度模型的参数量已经相当巨大，但如果换成浅层神经网络，其所需要的参数量要大出多个数量级才能达到相同的数据拟合效果，以至于很难实现。

深度模型之所以能减少参数的关键在于重复利用中间层的计算单元。以人脸识别为例，深度学习可以针对人脸图像的分层特征表达进行：最底层从原始像素开始学习滤波器，刻画局部的边缘和纹理特征；中层滤波器通过将各种边缘滤波器进行组合，描述不同类型的人脸器官；最高层描述的是整个人脸的全局特征。

深度学习提供的是分布式的特征表示。在最高的隐含层，每个神经元代表一个属性分类器（如图 1(b) 所示），例如性别、人种和头发颜色等。每个神经元将图像空间一分为二，N 个神经元的组合就可以表达 2^N 个局部区域，而用浅层模型表达这些区域的划分至少需要 2^N 个模板。由此可以看出，深度模型的表达能力更强，效率更高。

提取全局特征和上下文信息的能力

深度模型具有强大的学习能力和高效的特征表达能力，更重要的优点是从像素级原始数据到抽象的语义概念逐层提取信息，这使得它在提取图像的全局特征和上下文信息方面具有突出的优势，为解决传统的计算机视觉问题（如图像分割和关键点检测）带来了新的思路。

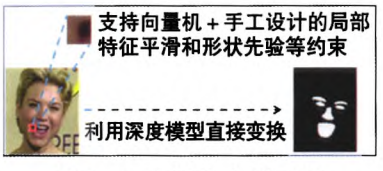

图2 人脸图像分割示例

以人脸的图像分割为例（如图 2 所示），为了预测每个像素属于哪个脸部器官（眼睛、鼻子、嘴），通常的做法是在该像素周围取一个小区域，提取纹理特征（例如局部二值模式），再基于该特征利用支持向量机等浅层模型分类。因为局部区域包含的信息量有限，往往产生分类错误，因此要对分割后的图像加入平滑和形状先验等约束。

人眼即使在存在局部遮挡的情况下也可以根据脸部其他区域的信息估计被遮挡部分的标注。由此可知全局和上下文信息对于局部的判断是非常重要的，而这些信息在基于局部特征的方法中在最开始阶段就丢失了。理想情况下，模型应该将整幅图像作为输入，直接预测整幅分割图。图像分割可以被看做一个高维数据转换的问题来解决。这样不但利用到了上下文信息，模型在高维数据转换过程中也隐式地加入了形状先验。但是由于整幅图像内容过于复杂，浅层模型很难有效地捕捉全局特征。而深度学习的出现使这一思路成为可能，在人脸分割[11]、人体分割[12]、人脸图像配准[13]和人体姿态估计等各个方面都取得了成功[14]。

联合深度学习

一些研究计算机视觉的学者将深度学习模型视为黑盒子，这种看法是不全面的。传统计算机视觉系统和深度学习模型存在着密切的联系，利用这种联系可以

提出新的深度模型和训练方法。用于行人检测的联合深度学习[15]就是一个成功的例子。一个计算机视觉系统包含若干个关键的组成模块。例如，一个行人检测器包括特征提取、部件检测器、部件几何形变建模、部件遮挡推理、分类器等模块。在联合深度学习中[15]，深度模型的各个层和视觉系统的各个模块可以建立对应关系。如果视觉系统中的关键模块在现有深度学习的模型中没有与之对应的层，则它们可以启发我们提出新的深度模型。例如，大量物体检测的研究工作表明，对物体部件的几何形变建模可以有效提高检测率，但是在常用的深度模型中没有与之相对应的层，因此联合深度学习[15]及其后续的工作[16]都提出了新的形变层和形变池化层[1]来实现这一功能。

从训练方式上看，计算机视觉系统的各个模块是逐一训练或手工设计的。在深度模型的预训练阶段[2]，各个层也是逐一训练的。如果我们能够建立计算机视觉系统和深度模型之间的对应关系，那么在视觉研究中积累的经验就可以对深度模型的预训练提供指导。这样预训练后得到的模型就可以达到与传统计算机视觉系统可比的结果。在此基础上，深度学习还会利用反向传播对所有层进行联合优化，使它们之间的相互协作达到最优，从而使整个网络的性能得到重大提升。

深度学习在物体识别中的应用

ImageNet图像分类

深度学习在物体识别中最重要的进展体现在ImageNet ILSVRC[3]挑战中的图像分类任务。传统计算机视觉方法在此测试集上最低的错误率是26.172%。2012年，欣顿的研究小组利用卷积网络把错误率降到了15.315%。此网络结构被称为Alex Net[3]，与传统的卷积网络相比，它有三点与众不同之处：首先，Alex Net采用了dropout的训练策略，在训练过程中将输入层和中间层的一些神经元随机置零。这模拟了噪音对输入数据的各种干扰使一些神经元对一些视觉模式产生漏检的情况。Dropout使训练过程收敛得更慢，但得到的网络模型更加鲁棒。其次，Alex Net采用整流线型单元作为非线性的激发函数。这不仅大大降低了计算的复杂度，而且使神经元的输出具有稀疏的特征，对各种干扰更加鲁棒。第三，Alex Net通过对训练样本镜像映射和加入随机平移扰动，产生了更多的训练样本，减少了过拟合。

在ImageNet ILSVRC 2013比赛中，排名前20的小组使用的都是深度学习技术。获胜者是纽约大学罗伯·费格斯(Rob Fergus)的研究小组，所采用的深度模型是卷积网络，并对网络结构作了进一步优化，错误率为11.197%，其模型称作Clarifai[17]。

在ILSVRC 2014比赛中，获胜者GooLeNet[18]将错误率降到了6.656%。GooLeNet突出的特点是大大增加了卷积网络的深度，超过了20层，这在此之前是不可想象的。很深的网络结构给预测误差的反向传播带了困难，这是因为预测误差是从最顶层传到底层的，传到底层的误差很小，难以驱动底层参数的更新。GooLeNet采取的策略是将监督信号直接加到多个中间层，这意味着中间层和底层的特征表示也要能够对训练数据进行准确分类。如何有效地训练很深的网络模型仍是未来研究的一个重要课题。

虽然深度学习在ImageNet上取得了巨大成功，但是很多应用的训练集是较小的，在这种情况下，如何应用深度学习呢？有三种方法可供参考：(1)可以将ImageNet上训练得到的模型作为起点，利用目标训练集和反

[1] 池化操作是在特征分布图的一个局部区域内取最大值或平均值传到神经网络下一层的特征分布图。经过池化操作，输出的特征分布图对局部形变具有更好的鲁棒性。

[2] 预训练是对神经网络的各个层次逐一优化，从而使网络参数达到一个好的初始点。人们通常在预训练之后，利用反向传播对所有层次的网络参数进行联合优化，进一步提高网络的性能。

[3] ILSVRC，大规模视觉识别挑战赛(Large Scale Visual Recognition Challenge)。

向传播对其进行继续训练，将模型适应到特定的应用[10]。此时ImageNet起到预训练的作用。(2) 如果目标训练集不够大，可以将底层的网络参数固定，沿用ImageNet上的训练集结果，只对上层进行更新。这是因为底层的网络参数是最难更新的，而从ImageNet学习得到的底层滤波器往往描述了各种不同的局部边缘和纹理信息，而这些滤波器对一般的图像有较好的普适性。(3) 直接采用ImageNet上训练得到的模型，把最高的隐含层的输出作为特征表达，代替常用的手工设计的特征[19, 20]。

人脸识别

深度学习在物体识别上的另一个重要突破是人脸识别。人脸识别的最大挑战是如何区分由于光线、姿态和表情等因素引起的类内变化和由于身份不同产生的类间变化。这两种变化的分布是非线性的，且极为复杂，传统的线性模型无法将它们有效区分开。深度学习的目的是通过多层的非线性变换得到新的特征表示。这些新特征须尽可能多地去掉类内变化，而保留类间变化。

人脸识别包括人脸确认和人脸辨识两种任务。人脸确认是判断两张人脸照片是否属于同一个人，属于二分类问题，随机猜的正确率是50%。人脸辨识是将一张人脸图像分为N个类别之一，类别是由人脸的身份定义的。这是个多分类问题，更具挑战性，其难度随着类别数的增多而增大，随机猜的正确率是1/N。两种任务都可以通过深度模型学习人脸的特征表达。

表2 不同深度学习算法在LFW确认任务上的识别率

方法	准确率	训练集人脸图像数
文献[21]	92.52%	87628
DeepFace[23]	97.35%	700万
DeepID[22]	97.45%	202599
DeepID2[24]	99.15%	202599
DeepID2+[8]	99.47%	45万

2013年，文献[21]采用人脸确认任务作为监督信号，利用卷积网络学习人脸特征，在LFW上取得了92.52%的识别率。这一结果虽然与后续的深度学习方法相比较低，但也超过了大多数非深度学习算法。由于人脸确认是一个二分类问题，用它学习人脸特征的效率比较低，容易在训练集上发生过拟合。而人脸辨识是一个更具挑战性的多分类问题，不容易发生过拟合，更适合通过深度模型学习人脸特征。另一方面，在人脸确认中，每一对训练样本被人工标注成两类中的一类，所含信息量较少。而在人脸辨识中，每个训练样本都被人工标注成N类之一，信息量大。

在2014年的IEEE国际计算机视觉与模式识别会议(IEEE Conference on Computer Vision and Pattern Recognition, CVPR)上，DeepID[22]和DeepFace[23]都采用人脸辨识作为监督信号，在LFW上分别取得了97.45%和97.35%的识别率(见表2)。他们利用卷积网络预测N维标注向量，将最高的隐含层作为人脸特征。这一层在训练过程中要区分大量的人脸类别(例如在DeepID中区分1000个类别的人脸)，因此包含了丰富的类间变化的信息，有很强的泛化能力。虽然训练中采用的是人脸辨识任务，但得到的特征可以应用到人脸确认任务中，以及识别训练集中是否有新人。例如，LFW上用于测试的任务是人脸确认任务，不同于训练中的人脸辨识任务；DeepID[21]和DeepFace[22]的训练集与LFW测试集的人物身份是不重合的。

通过人脸辨识任务学习得到的人脸特征包含较多的类内变化。DeepID2[24]联合使用人脸确认和人脸辨识作为监督信号，得到的人脸特征在保持类间变化的同时使类内变化最小化，从而将LFW上的人脸识别率提高到99.15%。DeepID2利用Titan GPU提取一幅人脸图像的特征只需要35毫秒，而且可以离线进行。经过主元分析(Principal Component Analysis, PCA)压缩最终得到80维的特征向量，可以用于快速人脸在线比对。在后续工作中，DeepID2[8]通过扩展网络结构，增加训练数据，以及在每一层都加入监督信息，在LFW达到了99.47%的识别率。

一些人认为深度学习的成功是由于用具有大量参数的复杂模型去拟合数据集，其实远非如此

简单。例如 DeepID2+ 的成功还在于其所具有的很多重要有趣的特征[8]：它最上层的神经元响应是中度稀疏的，对人脸身份和各种人脸属性具有很强的选择性，对局部遮挡有很强的鲁棒性。在以往的研究中，为了得到这些属性，我们往往需要对模型加入各种显示的约束。而 DeepID2+ 通过大规模学习自动拥有了这些属性，其背后的理论分析值得未来进一步研究。

深度学习在物体检测中的应用

物体检测是比物体识别更难的任务。一幅图像中可能包含属于不同类别的多个物体，物体检测需要确定每个物体的位置和类别。2013 年，ImageNet ILSVRC 比赛的组织者增加了物体检测的任务，要求在 4 万张互联网图片中检测 200 类物体。比赛获胜者使用的是手动设计的特征，平均物体检测率 (mean Averaged Precision, mAP) 只有 22.581%。在 ILSVRC 2014 中，深度学习将平均物体检测率提高到了 43.933%。较有影响力的工作包括 RCNN[10]、Overfeat[25]、GoogLeNet[18]、DeepID-Net[16]、network in network[26]、VGG[27] 和 spatial pyramid pooling in deep CNN[28]。RCNN[10] 首次提出了被广泛采用的基于深度学习的物体检测流程，并首先采用非深度学习方法（例如 selective search[29]）提出候选区域，利用深度卷积网络从候选区域提取特征，然后利用支持向量机等线性分类器基于特征将区域分为物体和背景。DeepID-Net[16] 进一步完善了这一流程，使得检测率有了大幅提升，并且对每一个环节的贡献做了详细的实验分析。深度卷积网络结构的设计也至关重要，如果一个网络结构能够提高图像分类任务的准确性，通常也能显著提升物体检测器的性能。

深度学习的成功还体现在行人检测上。在最大的行人检测测试集 (Caltech[30]) 上，广泛采用的方向梯度直方图 (Histogram of Oriented Gradient, HOG) 特征和可变形部件模型 [31] 的平均误检率是 68%。目前基于深度学习检测的最好结果是 20.86%[32]。在最新的研究进展中，很多被证明行之有效的物体检测都用到了深度学习。例如，联合深度学习 [15] 提出了形变层，对物体部件间的几何形变进行建模；多阶段深度学习 [33] 可以模拟物体检测中常用的级联分类器；可切换深度网络 [34] 可以表达物体各个部件的混合模型；文献 [35] 通过迁移学习将一个深度模型行人检测器自适应到一个目标场景。

深度学习用于视频分析

深度学习在视频分类上的应用还处于起步阶段，未来还有很多工作要做。描述视频的静态图像特征可以采用从 ImageNet 上学习得到的深度模型，难点是如何描述动态特征。以往的视觉研究方法对动态特征的描述往往依赖于光流估计、对关键点的跟踪和动态纹理。如何将这些信息体现在深度模型中是个难点。最直接的做法是将视频视为三维图像，直接应用卷积网络 [36] 在每一层学习三维滤波器。但是这一思路显然没有考虑到时间维和空间维的差异性。另外一种简单但更加有效的思路是，通过预处理计算光流场或其他动态特征的空间场分布，作为卷积网络的一个输入通道 [37-39]。也有研究工作利用深度编码器 (deep autoencoder) 以非线性的方式提取动态纹理 [38]。在最新的研究工作中 [41]，长短时记忆网络 (Long Short-Term Memory, LSTM) 受到广泛关注，它可以捕捉长期依赖性，对视频中复杂的动态建模。

未来发展的展望

深度学习在图像识别中的应用方兴未艾，未来有着巨大的发展空间。

在物体识别和物体检测研究的一个趋势是使用更大更深的网络结构。在 ILSVRC 2012 中，Alex Net 只包含了 5 个卷积层和两个全连接层。而在 ILSVRC2014 中，GooLeNet 和 VGG 使用的网络结构都超过了 20 层。更深的网络结构使得反向传播更加困难。与此同时，训练数据的

规模也在迅速变大。这迫切需要研究新的算法和开发新的并行计算系统来更加有效地利用大数据训练更大更深的模型。

与图像识别相比，深度学习在视频分类中的应用还远未成熟。从 ImageNet 训练得到的图像特征可以直接有效地应用到各种与图像相关的识别任务（例如图像分类、图像检索、物体检测和图像分割等）和其他不同的图像测试集中，具有良好的泛化性能。但是深度学习至今还没有得到类似的可用于视频分析的特征。要达到这个目的，不但要建立大规模的训练数据集（文献 [42] 最新建立了包含 100 万个 YouTube 视频的数据库），还需要研究适用于视频分析的新的深度模型。训练用于视频分析的深度模型的计算量也会大大增加。

在与图像和视频相关的应用中，深度模型的输出预测（例如分割图或物体检测框）往往具有空间和时间上的相关性。因此研究具有结构性输出的深度模型也是一个重点。

虽然神经网络的目的在于解决一般意义上的机器学习问题，但领域知识对深度模型的设计也起着重要的作用。在与图像和视频相关的应用中，最成功的是深度卷积网络，其设计正是利用了图像的特殊结构。其中最重要的两个操作——卷积和池化都来自与图像相关的领域知识。如何通过研究领域知识，在深度模型中引入新的有效的操作和层，对于提高图像和视频识别的性能有着重要意义。例如，池化层带来了局部的平移不变性，文献 [16] 中提出的形变池化层在此基础上更好地描述了物体各个部分的几何形变。在未来研究中，可以将其进一步扩展，从而取得旋转不变性、尺度不变性和对遮挡的鲁棒性。

通过研究深度模型和传统计算机视觉系统之间的关系，不但可以帮助我们理解深度学习成功的原因，还可以启发新的模型和训练方法。联合深度学习 [15] 和多阶段深度学习 [33] 未来还有更多的工作要做。

虽然深度学习在实践中取得了巨大成功，而且通过大数据训练得到的深度模型体现出的特性（例如稀疏性、选择性和对遮挡的鲁棒性 [8]）引人注目，但其背后的理论分析还有许多工作需要完成。例如，何时收敛？如何取得较好的局部极小点？每一层变换取得了哪些对识别有益的不变性，又损失了哪些信息？最近马拉特 (Mallat) 利用小波对深层网络结构进行了量化分析 [43]，这是在此方向上的重要探索。

结语

深度模型并非黑盒子，它与传统的计算机视觉系统有着密切的联系，神经网络的各个层通过联合学习、整体优化，使得性能得到大幅提升。与图像识别相关的各种应用也在推动深度学习在网络结构、层的设计和训练方法各个方面的快速发展。可以预见在未来数年内，深度学习将会在理论、算法和应用各方面进入高速发展时期。■

王晓刚
香港中文大学助理教授。主要研究方向为计算机视觉、深度学习、群体视频监控、物体检测和人脸识别等。xgwang@ee.cuhk.edu.hk

参考文献

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. *Nature*, 1986; 323(99):533~536.

[2] J. Deng, W. Dong, R. Socher, and et al.. Imagenet: A large-scale hierarchical image database. In IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2009.

[3] A. Krizhevsky, L. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. Neural Information Processing Systems, 2012.

[4] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miler. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst, 2007.

[5] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In IEEE Int'l Conf. Computer Vision, 2009.

更多参考文献：www.ccf.org.cn/cccf