



基于深度学习理论的 纹身图像识别与检测研究

许庆勇 著

The Research of
Tattoo Recognition
and Detection Based
on Deep Learning



华中科技大学出版社
<http://www.hustp.com>

基于深度学习理论的纹身图像识别 与检测研究

The Research of Tattoo Recognition and
Detection Based on Deep Learning

许庆勇 著



华中科技大学出版社
中国·武汉

内 容 提 要

本书针对纹身图像的局部性、内容的复杂性、纹理的清晰性、颜色的单一性、图案 Logo 的标志性、大小形状的多样性等特点,运用深度学习理论,从纹身的特点、研究理论、纹身识别与检测研究三方面进行系统、科学的研究。纹身图像识别与检测研究是最近几年才引起学者和有关人员重视的,其研究发展也是一个不断完善的过程,希望本书能对相关领域的研究起到一定的作用。

图书在版编目(CIP)数据

基于深度学习理论的纹身图像识别与检测研究/许庆勇著. —武汉:华中科技大学出版社, 2018. 12

ISBN 978-7-5680-4720-3

I. ①基… II. ①许… III. ①人工智能-算法-应用-文身-图像识别-研究 IV. ①TP391.41

中国版本图书馆 CIP 数据核字(2018)第 275273 号

基于深度学习理论的纹身图像识别与检测研究

许庆勇 著

Jiyu Shendu Xuexi Lilun de Wenshen Tuxiang Shibie yu Jiance Yanjiu

策划编辑:王 乾 李 欢

责任编辑:何 欢

封面设计:刘 婷

责任校对:阮 敏

责任监印:徐 露

出版发行:华中科技大学出版社(中国·武汉)

电话:(027)81321913

武汉市东湖新技术开发区华工科技园

邮编:430223

录 排:华中科技大学惠友文印中心

印 刷:北京虎彩文化传播有限公司

开 本:710mm×1000mm 1/16

印 张:9.25 插页:1

字 数:157 千字

版 次:2018 年 12 月第 1 版第 1 次印刷

定 价:39.80 元



华中出版

本书若有印装质量问题,请向出版社营销中心调换
全国免费服务热线:400-6679-118 竭诚为您服务
版权所有 侵权必究

前言

随着图像拍摄设备、智能手机和互联网技术的发展，纹身图像的采集、传播变得越来越容易。而网络技术的发展，特别是无线高速网络的发展(如 4G、5G 技术)，为网络图像分享提供了便捷的通道，出现了一些专业化的网络数据库和图像分享网站(如 www.flickr.com、www.sohuphoto.com、image.baidu.com、image.google.com 等)。伴随着突发事件的不断发生，纹身同其他生物特征一样，成为对犯罪嫌疑人识别的有力证据。如何对纹身图像进行检测和语义解读，已经引起公共安全部门的重视。

在 19 世纪 60 年代到 80 年代，俄国学者阿尔卡季布维尼科夫游走俄罗斯各地的惩教中心，在惩教中心寻找带有纹身的罪犯并拍下罪犯身上的纹身图案。19 世纪末 20 世纪初，波兰监狱管理局会对监狱每一位罪犯进行纹身，以更好地进行识别。当这些罪犯死亡后，罪犯身上的纹身图案会经由专门人员切割下来并保存，以便日后分析不同罪犯之间的关系。在我国，纹身也有着悠久的发展历史，但是大部分时间里，纹身都受到一定的歧视。例如纹身者被看作与罪犯、流氓、黑社会有关。早在中国封建时代，纹身作为一种社会符号，在罪犯身体的明显部位，如在罪犯的面部额下刺写“流放”、“充军”等字样来进行罪犯的标识和识别。由于受到传统观念(身体发肤，受之父母)的影响，普遍认为纹身是破坏自己的身体，是对父母的大不敬。

通过对各种纹身图案的解读，并借助纹身图案的油墨痕迹可查明大量犯罪嫌疑人犯下的各种罪行，例如通过不同的纹身图案(如玫瑰、蛇等)的解读而了解到这些罪犯的行为取向，以找到他们的同伙及其犯罪证据。随着公共安全事件的频繁出现，人们对纹身研究的关注度越来越高。

纹身同伤疤、斑、痣等一样成为人体生物特征的重要组成部分。同时，由于它提供了足以证明个人身份的信息，而受到越来越多的重视。

在一些黑社会团体中，纹身作为其入会的考验，也普遍被采用。一些公共安全管理部门开始通过纹身特征建立某种关联，以根据纹身进行分类来发现犯罪嫌疑人。对纹身的分类、识别、检测逐渐被学者和相关部门重视。本研究正是在此背景下开展的，运用深度学习理论和方法，针对纹身图像的局部性、内容的复杂性、纹理的清晰性、颜色的单一性、图案 Logo 的标志性、大小形状的多样性等特点进行研究。我们通过对不同方法和不同数据集的对比研究，取得了一定的研究成果。希望引起学者们对纹身研究的关注和更深入的研究。

许庆勇

2018 年 7 月

目 录

CONTENTS

第一章 绪论	1
一、引言	3
二、深度学习	5
三、纹身图像	12
四、纹身图像识别与检测	22
第二章 纹身图像处理理论基础	27
一、图像底层特征	29
二、词包模型	36
三、空间金字塔概述	37
四、深度置信网络概述	38
五、卷积神经网络	42
第三章 深度学习算法在纹身图像检测中的比较研究	47
一、深度学习算法	49
二、对比实验与分析	56
第四章 基于多特征融合的 DBN 纹身图像识别与检测	63
一、引言	65
二、基于多特征融合的 DBN 纹身图像检测改进算法	66
三、基于视觉词包的 DBN 纹身图像检测改进算法(BOVW-DBN)	68

四、基于空间金字塔的 DBN 纹身图像检测改进算法(SP-DBN)	74
五、基于 DBN 改进算法在 Caltech 101 分类中的应用	78
六、基于 DBN 改进算法在纹身图像检测中的应用	82
第五章 基于全连接层的 CNN 纹身图像检测	93
一、CNN 基本结构	96
二、基于 CNN 的改进算法	97
三、基于三通道 R-CNN 的纹身检测(CFT Faster R-CNN)	118
主要参考文献	130
后记	143

Chapter

1

.....

第一章 绪论



一、引言

互联网的快速发展,特别是 4G 移动通信技术和未来 5G 移动通信技术的发展,使大数据、高质量的图像信息传播效率越来越高。同时,网络技术、计算机技术、通信技术的高速发展,也使得信息表达形式从传统单一的文本表达逐渐向集文本、图像、音频、视频等形式于一体的多元化方向发展。这种多元化主要表现在信息的表现形式,如文本信息、图像信息、视频信息和音频信息。在这些表现形式中,图像因具有较强的综合性和直观性,而被人们广泛地获取和利用。据统计,人们日常生活的信息中有约 80% 信息来自视频,而在视觉信息中最直观的表现就是图像。随着图像拍摄设备的发展,如各种专业相机和便携式的拍摄设备(特别是当代智能手机),使得图像获取变得越来越容易;而网络技术的发展,特别是无线高速网络的发展(如 4G、5G 技术),为网络图像分享提供了便捷的通道,出现了一些专业化的网络数据库和图像分享网站(如 www.flickr.com、www.sohuphoto.com、image.baidu.com、image.google.com 等)。同时各种专业图像数据库在不断增加、图像集的规模也不断增大,这些数据库基本上覆盖了安全、军事、医疗和社会生活等各个方面,如美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)系列包含了 50 多个种类的数据库。

数字图像作为各种信息的重要的载体,以其直观、形象、信息量大等特点影响着人们生活的方方面面,成为人们获取信息的一个重要途径。在这些应用中,一个比较典型的例子就是纹身图像。纹身(Tattoo)是以一种特殊的方法(一般是用带有墨的针)刺入人的皮肤层,然后在皮肤上人为地制造出一些对自己有特殊意义的字、词、标志、吉祥物等图案。纹身一旦刺在皮肤上,将永久存在。纹身作为身体的一种标记,常与伤疤、斑、痣等特征作为身体的外在特征一起用于图像检测和识别。由于纹身提供了足以证明个人身份的标志,而被越来越多的组织和个人重视,比如犯罪团伙、证人、黑社会团体等。在一些黑社会团体中,纹身作为其入会的考验,也普遍被采用。在 20 世纪初,各街帮和摩托车帮成员也通过纹身作为其成员的标识。公元前 2000 年,在埃及木乃伊身上发现了多种精美的纹身图案;19 世纪晚期,纹身也曾流行于英国上层社会中,有大量的男男女女

进行纹身^①。

19 世纪末 20 世纪初,波兰监狱管理局会对每一位监狱罪犯进行纹身,以更好地进行识别。当这些罪犯死亡后,罪犯身上的纹身图案会经过专门的人员进行切割并保存下来,以便以后分析不同罪犯之间的关系。

根据统计数据显示,在美国,人们每年在纹身上的消费高达 16 亿美元;在 18~25 岁的成年人中,有 36% 的人至少有一个纹身;在 26~40 岁的成年人中,有 40% 的人至少有一个纹身,而且这一数据还在逐年增加^②。19 世纪,美国监狱管理部门在释放罪犯时,要对其进行纹身;俄国和德国的管理部门要在西伯利亚的监狱和集中营的在押人员身上进行纹身,以便以后进行识别和发现。2014 年,美国国家标准与技术研究院开始对纹身进行研究,NIST 这项标准把纹身共分为 8 个大类和 70 个小类。美国联邦调查局(FBI)根据 2014 年美国国家标准与技术研究院对纹身的分类标准对嫌疑人的纹身图案进行了分类。通过长时间积累,美国联邦调查局建立了一个纹身数据库,以此数据库为基础,开展下一代身份识别。美国联邦调查局以“囚犯纹身”数据库为基础开展了相关研究,这项研究将通过相同主题的纹身,建立人与人之间的联系或者根据人们的纹身对他们进行分类以便发现或查找犯罪嫌疑人。但是很多人会通过不同的方式来描述同一个图案,或通过不同图案来表达同一个含义。这样使得纹身的分类和表达变得更加难以理解。不同的人对同一个纹身图案会有不同的理解或者会想到不同的含义,这完全依赖于个人的主观色彩。例如,来自 NIST 的学者 Mei Ngan^③ 在研究中发现,底特律老虎队使用的字母“D”的徽章就经常被人误解,有人认为这是团队的象征,也有人认为它就是一个字母,或者一种抽象的设计^④。但是对于一个团体来讲,同一个纹身图案表达的含义总是明确、无异议的。

在我国,纹身也有着悠久的发展历史。但是在很长时间里,纹身都受到一定的歧视,例如纹身者被看作与罪犯、流氓、黑社会有关。早在中国封建时代,纹身

① 纹身百科:纹身为什么会进入美国主流社会,<http://wenshen.baike.com/article-92738.html>.

② <http://www.statisticbrain.com/tattoo-statistics/>.

③ Ngan M, Grother P. Tattoo recognition technology-challenge (Tatt-C): an open tattoo database for developing tattoo recognition research[C]//Identity, Security and Behavior Analysis (ISBA), 2015 IEEE International Conference on. IEEE, 2015.

④ FBI 想通过纹身来识别罪犯事情没那么简单,<http://roll.sohu.com/20150907/n420551277.shtml>.

作为一种社会符号,在犯人身体的明显部位,如在犯人的额下刺入“流放”、“充军”或其他罪名来进行犯人的标识和识别。由于受到传统观念(身体发肤,受之父母)的影响,人们普遍认为纹身是破坏自己的身体,是对自己父母的大不敬。由于以前在中国纹身的成员当中,以社会的“混混”、流氓人员居多,所以纹身者的数量一直不多。但随着人们观念的变化,受名人和体育明星等的影响,纹身越来越多地被社会认同。特别是“00后”的年轻一代认为,纹身是塑造自己身份、铭记自己生活经历的一种方式。根据统计数据显示,43%纹身者认为纹身对其有特殊的意义。^① 总之随着时代的发展,纹身者的数量也在逐渐增加。

随着时间的推移,纹身者的数量也急剧增加。例如,Flickr上已有数百万纹身图像,这些纹身图案虽然来自不同的团体和个人,但是对纹身图案的区分、检测与分类却有着极其重要的意义;NIST的纹身数据库,在纹身图像的检测与分类方面有2349张图像,其中纹身图像1349张,非纹身图像1000张;CEA、Compass、MITRE、MorphoTrak等以这个数据集为基础进行了纹身图像检测的相关研究,取得了一定的成果,这个数据集相对于其他类别的图像数据集来讲,显得相对较小,因此有很多学者开始对Flickr上的纹身图案进行研究。通过对Flickr上的纹身图案研究进而找出不同犯罪嫌疑人身上纹身图案的关系。

如何对纹身图像进行快速分类,并高效查询出与之相似的纹身图案,进行语义的解读,以发现其正常或不正常的行为,并为相关部门和人员提供有力的证据,越来越多地被安全部门重视,例如美国NIST机构。同时,对纹身图像进行正确标注和分类等都成为计算机图像领域研究的重要内容。

二、深度学习

深度学习(Deep Learning)是在机器学习的基础上发展起来的。机器学习(Machine Learning, ML)是一门涉及多学科、多领域的交叉学科,主要包括概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。^② 机器学习是人工智能的核心,在某种程度上,它使计算机智能化。

^① <http://www.statisticbrain.com/tattoo-statistics/>.

^② <https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/217599?fr=aladdin>.

深度学习的目标是使机器学习能够更加接近其最初的目标——人工智能^①。从2006年Hinton提出深度置信网络结构开始,深度学习作为机器学习领域中对模式(音频、图像、文本等)进行建模的一种方法已经成为机器学习研究的一个新领域。

最近几年,随着大家对深度学习研究的深入,大量学者致力于深度学习原理和应用的研究。在学术界主要体现在各种会议、期刊、高校研究组和企业应用上。在会议方面,著名的国际会议如IEEE Conference on CVPR、IEEE International Conference on ASSP、ACM Conference on Multimedia、SPIE Conference on Electronic Imaging等纷纷设专题交流最新的研究进展。全球深度学习峰会已连续举办10届,第11届于2017年4月在新加坡举办,对深度学习算法、企业应用等特点领域进行讨论。2017年中国电子学会在北京举办了全国深度学习应用大会。世界上各大著名的杂志和期刊纷纷设专刊介绍该领域最新的研究成果,各大著名杂志如IEEE Trans. on PAMI、IEEE Trans. on Image Processing、IEEE Trans. on Multimedia、IEEE Trans. on CSVT、Computer Vision and Image Understanding、International Journal of Computer Vision and Pattern Recognition、Signal Processing、Computer Graphics and Interactive techniques、International Joint Conference on Artificial Intelligence、ACM International Conference on Multimedia、IEEE International Conference on Acoustics、Speech and Signal Processing等。高校团队如多伦多大学的Geoffrey Hinton研究组、斯坦福大学的Andrew Ng研究组、加拿大蒙特利尔大学的Yoshua Bengio研究组、纽约大学的Yann LeCun研究组等。企业团队有百度公司的Andrew Ng余凯团队、微软公司的邓力团队、Google公司的Geoffrey Hinton团队和阿里巴巴、科大讯飞以及中科院自动化所等公司或研究单位等。

深度学习作为新兴的机器学习算法,通过组合底层特征形成更加抽象的高层表示或特征,以发现数据的分布特征表示,例如通过形状、纹理和颜色等底层特征提取并进行融合,再通过深度学习算法,就可以实现纹身图案的检测。Hinton最早于2006年提出深度置信网络,后来LeCun等人又提出了卷积神经

① 深度学习概述:<http://blog.csdn.net/pingye9/article/details/50815348>.

网络,它利用有脑局部感受野的思想,通过权重共享以减少参数数量,达到提高BP训练性能。后来Bengio Y对深度学习进行了系统研究,包括常用的深度学习方法(DBN、CNN、SC、AE等);Bengio Y认为深度学习通过增加网络深度,学习到一种深层的非线性网络结构,实现对输入数据的分布表示,并表现出强大的学习能力。通过非线性网络结构和深度学习的网络学习,可以有效地实现对纹身的数据表示。

在应用方面,深度学习在图像分类、语音识别、自然语言处理等领域取得了较好的成果。2012年11月,微软在一次活动上展示了一个通过深度学习实现的同声传译系统,中英文翻译和语音合成效果非常好。在2013年的百度年会上,李彦宏宣布要成立百度研究院,其中第一个成立的就是深度学习研究所(Institute of Deep Learning,IDL)。2017年,发改委批复由百度牵头组建深度学习国家级实验室,重点对深度学习技术、计算机视觉感知技术、计算机听觉技术、生物特征识别技术、新型人机交互技术、标准化服务、深度学习知识产权七大方向进行研究。百度语音技术入选MIT(麻省理工学院)科技评论杂志评出的“2016年全球十大突破技术”,百度也被该杂志评选为全球最具智慧的公司排行榜第二位。2014年3月,Fan等人使用空间金字塔卷积神经网络在人脸识别领域中达到了97.30%的正确率,该结果已接近于人类的识别水平。纹身图像检测作为图像研究的领域之一,也是深度学习的研究方向之一。

纵观各学者和研究团队的研究成果,一般认为深度学习是一组研究方法,对于传统的机器学习方法来讲的,其核心的深度学习方法主要包括四种,分别是深度置信网络(Deep Belief Network, DBN)、卷积神经网络(Convolutional Neural Network, CNN)、稀疏编码(Sparse Coding, SC)和自动编码器网络(Auto-encoders, AE)。这四种方法各有不同的特点,如深度置信网络最早被提出和应用,但其输入数据维度不宜过多,一般是几十维到数百维的特征作为输入;再如卷积神经网络,其本质是一种特征提取网络,在特征提取方面表现非常突出,最终需要通过SVM、SIGN等分类器进行分类,实现最终的目标。

1. 深度置信网络

深度置信网络作为几种主要的深度学习算法之一,它是受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)在深度架构上的推广,通过受限玻尔兹曼机模型堆叠而成的。深度置信网络也是目前研究和应用较为广泛的深度学习

架构之一。一般认为,深度置信网络是由多个隐层(隐层数大于2)组成的概率生成模型,并且可以有效地训练和表示非线性数据。每一层都从上一层获取相应的表示,并在观察数据和标签之间建立起一个联合概率分布。因此,深度置信网络的核心思想就是由下向上的每一层受限玻尔兹曼机对输入数据进行提取、抽象,尽可能保留重要信息的过程。

2006年,多伦多大学机器学习领域的著名教授 Hinton 在国际顶尖学术期刊《Science》上发表一篇文章,提出利用深度置信网络实现了数据降维及分类,引发了深度学习的浪潮。深度置信网络由多层受限玻尔兹曼机组成的深层神经网络结构,解决了传统的 BP 算法在训练多层神经网络过程中参数过多、训练速度过慢的难题。深度置信网络作为一种神经网络,本质就是通过函数逼近,学习数据的非线性表示,以提取数据中的深层信息。针对纹身图像的特点,纹身图像的检测与识别也是通过提取各种特征,包括底层颜色、纹理、形状特征、中层词包特征、SIFT 特征和金字塔特征,实现纹身图像的线性或非线性表示,以获取纹身图像的深层信息。

深度置信网络作为最重要的深度学习算法之一,目前已被成功应用于手写字体识别、语音识别等领域,取得了较好的效果。更多的应用,例如文本表示、音频事件分类和各种各样的可视化数据分析任务。Hinton 等人总结了用 DBN 代替高斯混合模型,与隐马尔科夫模型相结合,在语音识别领域取得了成果。Horster 和 Lienhart 用深度置信网络进行图像降维,并实现图像的检索。Torralba 等人使用深层架构实现了图像压缩,从而可以进行快速检索。Bazzani 等人使用改进后的深度置信网络来处理视频图像,进行物体的追踪与识别。Tang 等人将深度置信网络用于视觉感知领域,在只出现一次的人脸识别数据集上表现良好。在国内,如孙劲光提出数值属性的深度置信网络,并在 UCI(加州大学欧文分校)的多个数据集上进行对比验证,证明其有效性;陈丽萍等人把深度置信网络应用到说话人的语音信息提取中,通过 NIST SRE2008 数据库,验证其有效性;徐东辉等人把 DBN 应用到网络入侵检测中,通过 KDD99 数据集验证其有效性。

在 DBN 算法方面,众多学者以 DBN 为基础进行了改进,如 Salakhutdinov R、Hinton G E 提出了 DBN-rNCA 算法,利用 DBN 结构和近邻分析(Neighborhood Component Analysis, NCA)实现了降维;Zhong 等人证明了

DBN-rNCA 的有效性; Yan Liu 等人提出了差分 DBN 并证明了 DDBN (Discriminative Deep Belief Network, DDBN) 的有效性, DDBN 综合了 DBN 对无标签数据的区分力和 BP 算法对有标签数据的区分力; Honglak Lee 认为 DBN 没有很好地考虑输入数据的二维特点, 提出了 CDBN (Convolutional Deep Belief Network, CDBN); Zhong 提出了 BDBN (Bilinear Deep Belief Network, BDBN) 的分类算法; Shusen Zhou 基于 DBN 和模糊集理论提出了 FDBN (Fuzzy Deep Belief Networks, FDBN) 等等, 这些有针对性的改进, 在一定程度上提高了 DBN 算法在特定问题的性能。

在现有的对深度置信网络算法及应用研究中, 由于深度置信网络采用的是神经元输入的方式, 每一层的神经元的个数对深度置信网络的影响非常大, 因此各学者在对深度置信网络的应用方法上均采用像素比较低的手写字体数据集进行实验, 对多像素的(如 1000×1000)的图像处理和分析的相对比较少, 这也是深度置信网络存在的一个主要问题。

2. 卷积神经网络

卷积神经网络是另一种比较重要的深度学习算法, 作为深度学习方法的一种, 由于其优异的性能, 被学者广泛研究, 并成功应用到多个不同的领域。卷积神经网络最早由 LeCun 等人于 1990 年提出, 已成为当前语音分析和图像识别领域的研究热点。传统的观点认为, 反向传递神经网络(BP 神经网络)随着隐含层数量增加发生梯度弥散的问题, 即反向传递的误差信号在方向传递过程中越来越小, 对网络权值的影响也越来越小。1998 年, LeCun 在卷积神经网络的基础上进行了改进, 通过追叠卷积层和全连接层, 设计出了一个七层(三个卷积层、两个池化层和两个全连接层)的深度学习网络结构, 称为 LeNet5, 解决了这一问题, 并成功地应用到文字的识别中, 取得了较好的结果。2012 年, Hinton 的学生 Alex Krizhevsky 设计出一个新的网络结构, 称之为 CUDA-CONVNET, 这个网络结构包括五个卷积层、三个全连接层, 最后通过 softmax 函数输出 1000 值对 ImageNet 图库进行分类, 共计 1000 类图像, 最终取得了 TOP-5 为 83% 的正确率。随后引起了各学者对 CNN 的研究热潮, 各种改进卷积神经网络算法也被逐渐提出并被应用, 推动了卷积神经网络的进一步发展。对卷积神经网络的研究方法, 主要包括以下三个方面。

一是网络架构的设计。比较典型的卷积神经网络如 LeCun 的 LeNet5、

Alex 的 CUDA-CONVNET, 还有 CAFFE、GoogleNet、OverFeat、VGGNet 等。随着机器性能的提升, 这些网络架构的深度从最初的 5~8 层发展到几十层、几百层, 甚至更多。还有是对卷积核的大小、卷积步长、激活函数等方面进行设计。随着层数的增加, 网络的性能也在提升, 但提升的速度在减慢, 也就意味着要消耗更大的时间成本来换取较小的效果的提升, 但小幅度效果的提升在不同的应用方面的意义是不一样的。

二是网络的训练。卷积神经网络随着层数的增加, 网络参数在快速增加。因此在网络训练样本不足的情况下很容易出现过拟合现象, 这种情况影响了网络训练的收敛性能。为了减少这种过拟合, 学者提出了不同的研究方法, 如 Over-feat、Drop-out、Drop-connect 等。由于网络参数量大, 网络参数的初始化也成为一大问题, 常用的初始化方法主要有基于高斯分布的参数初始化、基于已训练好的模型的 Fine-tune 方式初始化、基于高斯分布的各层独立初始化等三类方法。同时为了减少参数的数量, 在卷积层通过权重共享的方式进行。同一层的不同卷积使用相同的权重值, 对性能有一定的提升。卷积神经网络通过采用权值共享的网络结构, 在一定程度上降低了网络模型的复杂度, 减少了权值的参数数量。这种方法在网络的输入是多维数据时, 表现得更为明显。

三是网络的应用。通过训练好的模型对样本进行预测, 在卷积神经网络不同的层次上抽取不同的特征信息, 最后通过全连接层的条件概率分布进行结果的分布判断, 表现出较强的判别能力和泛化能力。卷积神经网络利用图像的空间关系和权重共享, 减少了参数的学习, 提高 BP 的性能。在卷积神经网络中, 输入数据每通过一层, 就会获得观测数据的相应显著特征, 包括平移、缩放和旋转不变性等显著特征。然后通过逐层提取, 在高层就可以提取到图像的高级特征, 成功地应用到目标检测、图像分类、语音识别等领域, 如在目标检测方面, 先后有 RCNN、SPPNET、Fast-RCNN、Faster-RCNN、YOLO 等, 它们通过区域选择突破了传统卷积神经网络的限制。但是大量的候选区域对物体的检测速度产生了极大的影响。在其他方面, Khan 等把卷积神经网络应用于人脸检测和识别中; Levi 等利用 CNN 对人的性别和年龄进行分类与预测; Zhou 等利用 CNN 建立了图像识别和复杂场景识别之间的联系; Ji 利用 3D-CNN 实现了人的行为识别等。

但是, 对卷积神经网络的研究, 大家普遍关注的是网络的层次与结构、卷积