

文章编号: 2095-2163(2019)03-0294-04

中图分类号: TP183

文献标志码: A

图像视觉识别算法概述

任 萌

(西安石油大学 计算机学院, 西安 710065)

摘 要: 随着人工智能的快速发展,人们现在的生活也越来越智能化,各类人脸识别打卡,指纹检测技术越来越广泛地应用到人们的日常生活当中,这离不开图像视觉识别算法和相关技术的发展。其中,卷积神经网络对于视觉识别技术和方法的研究中已经占有绝对的优势。目前,许多研究学者尝试通过优化卷积神经网络算法的结构和性能来提高图像识别的精确度和识别速度,从而使得其可以广泛应用于数据挖掘的各个领域,并对其在实际生活的应用做出更大的贡献,本文将对 The Residual Network, DenseNet, GoogLeNet 三种图像视觉识别模型的训练过程、应用和特点进行研究,同时对图像视觉识别算法和技术在实际生活中的应用进行简要论述与展望,特别是在数据挖掘方面的应用。图像视觉识别技术是一种可以应用于各个领域,也可以剖析人类行为的热点研究课题。

关键词: 图像视觉识别; Residual Network; DenseNet; GoogLeNet; 卷积神经网络; 对比分析; 数据挖掘

Overview of image visual recognition algorithm

REN Meng

(School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China)

【Abstract】 With the rapid development of Artificial Intelligence, people's life is now more and more intelligent, all kinds of face recognition detection, fingerprint detection technology is more and more widely applied to people's daily life, which is inseparable from image visual recognition algorithm and the development of related technologies. Among them, Convolutional Neural Networks have an absolute advantage in visual recognition techniques and methods. At present, many researchers try to improve the accuracy and speed of image recognition by optimizing the structure and performance of Convolutional Neural Network algorithms, which makes it widely applicable to various fields of data mining and its application in real life. This paper introduces the training process, application and characteristics of the three image visual recognition models of The Residual Network, DenseNet, GoogLeNet, and briefly introduces and forecasts the application of image visual recognition algorithms and techniques in real life, especially in Data Mining applications. Image visual recognition technology is a research topic that can be applied to various fields and the analysis of human behavior.

【Key words】 image visual recognition; ResNet; DenseNet; GoogleNet; Convolutional Neural Network; comparative analysis; data mining

0 引 言

图像视觉识别(Image Visual Recognition)作为一种已经在现实生活中广泛应用的技术,近几年来,更是发展迅速。2014年, Szegedy 等人^[1]提出了 GoogLeNet,这是一种全新的深度学习结构,主要解决计算机硬件和稀疏结构之间的矛盾问题。基于卷积神经网络^[2](CNN)中类似于 LeNet-5 的标准结构而提出的。既引入稀疏性,又利用现有硬件针对密集矩阵的计算能力,设计了 Inception^[3-4]结构用于机器学习的分类任务中^[5],但参数太多,容易发生过拟合。2015年,微软的 He 等 4 名华人提出了 ResNet^[6-8](Residual Neural Network),使用 ResNet Unit 成功训练出了 152 层的神经网络结构,加快神

经网络的训练速度,模型的精度也有大幅提升。ResNet^[9]在图像识别^[10]、定位和检测任务中都取得了非常不错的效果。2017年, Huang 等人提出 DenseNet^[11],避开了 ResNet 的加深神经网络和 Inception 的加宽网络结构的方式,利用每一个输入层都会直接影响到后面所有输出层的思想,提升模型精度。

目前,图像视觉识别技术的应用已可见到很多,包括人脸检测^[12-13]、目标检测^[14-15]、以及百度、IBM 和 Google 推出的很多图像检测模块的产品都已经得到了广泛使用。通过辨识图片中的物体类别,都已经陆续应用到医学^[16]、车辆定损^[14,17]、农作物识别^[18]等等社会实际生活中,提高人们的生活品质。

但是,从数据挖掘^[19-21]的角度来看,图像识别

作者简介:任 萌(1995-),女,硕士研究生,主要研究方向:图像识别、可视化技术、数据挖掘。

收稿日期:2019-03-05

技术使得不仅仅可以从单一的文字来挖掘人们的行为和喜好,同样可以通过在线社交评论网站的图片来挖掘不同人的行为和喜好。这也是目前学界广受研究者关注的一个研究课题。

1 图像视觉识别算法

1.1 GoogLeNet

GoogLeNet 是 Google 研究出来的深度网络结构,通过提出 Inception 的网络结构来引出 2 个问题,对此可描述为:

- (1) 增加网络深度可以提高模型精确度。
- (2) 降维可以减少计算量。

其核心思想便是通过增加模型深度(层数)或者其宽度(神经元)来保证高质量的模型。如图 1 所示,整个 GoogLeNet 就是由多个 Inception 结构串联起来,其中利用 1×1 的卷积来进行升降维,并在多个尺寸上同时进行卷积再聚合。该图引自 Szegedy 等人的论文中。

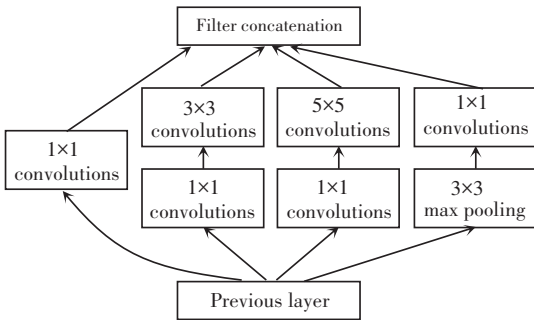


图 1 GoogLeNet 中的 Inception 结构

Fig. 1 Inception module in GoogLeNet

1.2 ResNet

ResNet 残差网络最初是由 He 等人^[7,22]提出来的,并由其进行扩展。通过采用了一种原则性方法,对每一层的输入做了一个参考,这种残差函数更容易优化,能使网络层数大大加深。研究中给出一个实例如图 2 所示。图 2(a)中,映射后可以拟合出任意目标映射 $H(x)$, 其中 x 表示网络层数,引出了一个 Residual net。图 2(b)拟合出 $F(x)$, 而且 $H(x) = F(x) + x$, 其中 $F(x)$ 表示残差映射,当 $H(x)$ 最优映射接近 identity 时,很容易捕捉到小的扰动。在计算机视觉里,随着网络深度的加深,特征的等级就会增高。有相关研究表明,网络的深度是取得良好成效的重要因素,然而梯度弥散则成为训练深层次网络的障碍,使其无法收敛。目前,已经相继推出了一些方法可以弥补这一缺陷,通过实现各层输入归一化,使得可以收敛的网络的深度提升为原来的 10

倍。只是,虽然达到了收敛的目的,但随着网络层数的增加会导致更大的误差,使得网络开始退化。

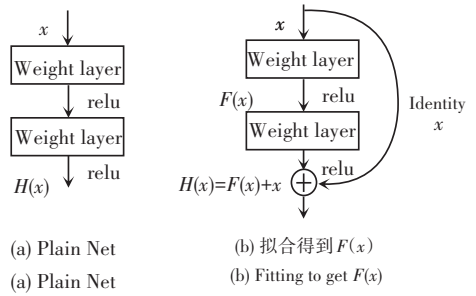


图 2 ResNet 实例

Fig. 2 ResNet example

若要获得更好的需要网络性能, $F(x) = H(x) - x$ 是 ResNet 学习的残差函数,如果 $F(x) = 0$, 那么此时的 $H(x) = x$ 为恒等映射,即没有引入额外的参数和计算复杂度,由此通过学习来找到对恒等映射的扰动会比重新学习一个映射函数要容易。基于前述处理,研究得到的就叫做残差块(Residual Block),而把这些残差块堆积起来就是残差网络了。这就是 ResNet 最主要的优点,直接将深层转化为前层,ShortCut(捷径)。研究可知,每个残差块可以表示为:

$$y_l = h(x_l) + F(x_l, W_l), \tag{1}$$

$$x_{l+1} = f(y_l), \tag{2}$$

其中, x_l 和 x_{l+1} 分别表示第 l 个残差块的输入和输出; F 是残差函数,表示学习到的残差;而 $h(x_l) = x_l$ 表示恒等映射; f 是激活函数; W 表示权值。

在式(1)、式(2)基础上,研究求得从前层 l 到深层 L 的学习特征,可写作如下数学形式:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, W_i). \tag{3}$$

然后通过计算梯度来消除残差块。

1.3 DenseNet

DenseNet 是以前馈的方式将各分层链接到每一层。主要由 DenseBlock 和 Transition 组成,可以缓解梯度消失问题,提高模型精度,并大幅减少参数数量。如此设计使得网络训练更加容易,每一层都可以利用原始的输入信息和损失函数的梯度,其设计结构如图 3 所示。这也更加有利于训练更深的网络。并且其稠密链接的正则化作用在少量的训练集中可以有效降低过拟合。同时,为了改善层之间的信息流,其设计研发是基于 ResNet 之上,相应的公式具体如下:

$$F(x_l) = H[x_0, x_1, x_2, \dots, x_l]. \tag{4}$$

其中, $H[x_0, x_1, x_2, \dots, x_l]$ 表示层中串联生成的特征和。这里仍需指出, 如果将该模型应用于大的数据集上时会损耗很大的硬件内存。因此建议将其应用于小数据集中。

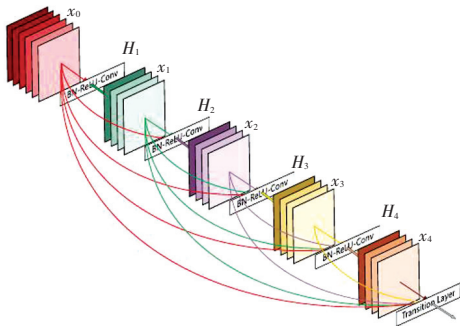


图3 DenseNet 结构

Fig. 3 DenseNet structure

2 应用

2.1 图像视觉识别算法的应用

自从图像视觉识别算法面世以来, 便引发了研究热潮, 无论在理论或是技术上均取得了长足进步, 其成果应用也已逐渐拓展至生活的方方面面。例如, 通用物体及场景识别和图像主体检测, 代替了人眼的辨识功能; 细粒度图像识别, 对图片中的动物、植物、菜品、logo 商标、车型等进行识别; 实现图片中的人流统计, 人体分析; 识别车辆外观受损部件及受损类型应用与车损检测中, 减少了人力劳动; 相册图片分类, 安装在手机端, 可以帮助人们整理相册, 节省手机内存等等。

特别地, 在数据挖掘推荐系统 (Recommendation System) 的有关研究、成果推广中, 比如, 像大众点评, 猫途鹰 (TripAdvisor) 等旅游或其它在线社交网络中, 若立足于消费者层面, 当其拟欲进行消费选择时, 如果想要根据自己的喜好, 屏蔽不想接触到的垃圾图片, 便可以通过图像视觉识别技术来过滤掉相关图片, 从而提高消费者的旅游和用餐体验。而立足于酒店或者相关管理人员的层面, 在推荐系统中, 不仅可以利用用户的评论和评分来帮助消费者进行推荐选择, 还可以根据用户以往在评论中发表的图片来匹配其在意向中最为青睐的酒店类型和消费体验。同样地, 用户出门在外通常会通过图片的方式来记录自己的生活, 那么, 还可以通过图像视觉识别技术来发现消费者的旅游轨迹, 从而在以后的旅游中更好地为消费者量身定制相关产品或服务, 使得消费者的生活更加便利与智能化。

有了图像视觉识别算法, 前文的图像视觉识别内容在社交网站的相关研究中实现起来就更加方便了。通过爬取在线社交网站的图片, 并识别其图像实体内容, 利用推荐系统和自然语言处理 (Natural Language Processing) 相关算法对文本信息进行处理, 来提取得到不同图片组的侧重点, 从而设计推出更有利于消费者和相关管理人员的网站, 这无论是对于消费者、网站管理员、或者酒店管理员来说, 均可从中得到好处。

2.2 算法对比

在前文中, 研究已经综合探讨了 3 种图像视觉识别算法。其中, DenseNet 是基于 GoogLeNet 和 ResNet 两个模型提出的。3 种图像视觉识别模型对比见表 1。由表 1 可以看出, GoogLeNet^[23] 最早进入公众视野, 但是其性能却逊色于后来的 2 个模型, 其参数较多, 并且容易产生过拟合, 若训练集有限, 这一特征将更加明显; 一旦训练的网络越大, 那么该模型的计算复杂度就会越大, 难以应用到实际场景中。针对不同的应用场景, 3 个模型的优势领域也不一样。其中, ResNet 和 DenseNet 更能受到广大研究学者的推崇与青睐, 其核心思想均是适用于 Highway Nets 小数据集的时候。因为数据集小, 就容易产生过拟合, 而 DenseNet 能更好地解决过拟合问题, 其泛化性能也更好。但是, 对于较大的数据集, ResNet 的性能要更胜一筹, 并且速度相对于 DenseNet 算法来说也更快。究其原因就在于, DenseNet 每一层都需要用到前面所有层的特征。这样看来, DenseNet 即便改善了算法精度, 解决了过拟合问题, 而且泛化性能有所提高, 但是该算法对硬件内存占用却会随着数据集增大而逐步提高, 对于硬件环境是非常不友好的。因此在设计中就需根据不同的应用场景已知的既有条件来选择研究使用的模型。

表 1 3 种图像视觉识别模型对比

Tab. 1 The contrast between three visual recognition models

模型	优点	缺点	适用
GoogLeNet	控制计算量和参数的同时, 有好的分类性能	参数太多, 容易过拟合, 精度不高	大数据集
ResNet	速度快	训练精度不高	大小数据集
DenseNet	很好解决过拟合问题, 泛化性能更好	数据集大时内存损耗高	小数据集

3 结束语

论文中综合探讨了 GoogLeNet、ResNet、DenseNet 三个模型的基本原理, 同时也对比了 3 个

模型的性能。从中不难发现,图像视觉识别算法正呈现出智能化态势,但是其中也存在着一定的缺陷与不足。尽管 DenseNet 参数、精度和结构以及调节方面有出色表现,但是其在大数据集的内存占用情况下却仍有待改善。而 ResNet 虽然赢得了业界的广泛认可,但在精度上却比不上 DenseNet 的训练精度。这也是后续工作中的一个重点研究方向。

目前,已有相当数量的研究人员正在尝试将图像视觉识别技术更广泛地应用在数据挖掘中,利用在线社交网站上的图片来发掘消费者的运行轨迹和其它社交习惯也不失为一个值得深入探索的项目研究课题。更多的研究实践均会证明,图像视觉识别技术在未来的科技进步中必将展示出较高的实用价值。

参考文献

[1] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[J]. arXiv preprint arXiv:1409.4842,2014.

[2] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA:IEEE, 2015:2625-2634.

[3] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional Residual networks[C]//IEEE International Conference on 3D Vision (3DV). Stanford, CA, USA:IEEE,2016:1-13.

[4] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167v2,2015.

[5] 张宏伟,张凌婕,李鹏飞,等.基于GoogLeNet的色织物花型分类[J].纺织科技进展,2017(7):33-35,52.

[6] LI Sihan, JIAO Jiantao, HAN Yanjun, et al. Demystifying ResNet[J]. arXiv preprint arXiv:1611.01186,2014.

[7] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[J]. arXiv preprint arXiv:1512.03385, 2015.

[8] XIE Saining, GIRSHICK R, DOLLÁR P. Aggregated residual

transformations for deep neural networks[J]. arXiv preprint arXiv:1611.05431,2016.

[9] TARG S, ALMEIDA D, LYMAN K. Resnet in Resnet: Generalizing residual architectures[J]. arXiv preprint arXiv:1603.08029,2016.

[10] WU Zifeng, SHEN Chunhua, HENGEL A V D. Wider or deeper: Revisiting the ResNet model for visual recognition[J]. Pattern Recognition, 2019,90:119-133.

[11] HUANG Gao, LIU Zhuang, van der MAATEN L, et al. Densely connected convolutional networks[J]. arXiv preprint arXiv:1608.06993,2016.

[12] 付学桐. 基于深度学习的人脸识别技术研究[J]. 通讯世界, 2019,26(2):299-300.

[13] 张裕平,龚晓峰,雒瑞森. 稀疏化的双向二维主成分人脸图像识别[J/OL]. 计算机工程:1-7[2019-02-28]. <http://kns.cnki.net/kcms/detail/31.1289.TP.20190122.1629.005.html>.

[14] 张长青,杨楠. 一种面向交通标志识别系统的图像识别算法[J/OL]. 电子科技:1-5[2018-12-20]. <http://kns.cnki.net/kcms/detail/61.1291.TN.20181220.0913.044.html>.

[15] RUSSAKOVSKY O, DENG Jia, SU Hao, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[16] 赖杰伟,陈韵岱,韩宝石,等. 基于DenseNet的心电数据自动诊断算法[J]. 南方医科大学学报,2019,39(1):69-75.

[17] 李晓飞. 基于深度学习的行人及骑车人车载图像识别方法[D]. 北京:清华大学,2016.

[18] 宋丽娟. 基于图像的农作物病害识别关键算法研究[D]. 西安:西北大学,2018.

[19] 凤祥云,卢辉斌. 基于数据挖掘的水下激光图像识别技术[J]. 激光杂志,2016,37(1):55-58.

[20] 杨帆,周华,王嫣. 基于数据挖掘的激光雷达图像识别技术[J]. 激光杂志,2016,37(7):39-42.

[21] 周诚诚,张代远. 利用图像识别技术过滤海量可疑钓鱼网站[J]. 计算机技术与发展,2012,22(11):246-249.

[22] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Identity mappings in deep residual networks[J]. arXiv preprint arXiv:1603.05027, 2016.

[23] KHAN R U, ZHANG Xiaosong, KUMAR R. Analysis of ResNet and GoogleNet models for malware detection[EB/OL].[2018-08-28]. <https://DOI.ORG/10.1007/S11416-018-0324-Z>.

[24] 黄奕晖,冯前进. 基于三维全卷积DenseNet的脑胶质瘤MRI分割[J]. 南方医科大学学报,2018,38(6):661-668.

(上接第293页)

[4] 李莉英,王其林. 浅谈通识教育背景下的高等数学教学[J]. 重庆文理学院学报(自然科学版),2011,30(6):97-99,102.

[5] 酆丽华,秦学礼. 智能手机操作系统的发展与分析[J]. 计算机教育,2014(6):70-74.

[6] 钮建伟,俞佳莉. 国内主流智能手机操作系统可用性评价[J]. 工业工程与管理,2014,19(4):104-108,134.

[7] 郭霖. 第一行代码—Android[M]. 2版. 北京:人民邮电出版社,2016.

[8] 肖中文. 影响中职数学应用能力培养的因素及对策[J]. 职业教育研究,2009(11):40-41.

[9] 杨浩宇. 数据库设计中E-R图向关系模式的转换[J]. 电脑编程技巧与维护,2012(2):50-51.

[10] 龙诗磊. UI设计与用户体验在产品中的重要性[J]. 西部皮革, 2016(20):79.

[11] 王翠萍. Android Studio 应用开发实战详解[M]. 北京:人民邮电出版社,2017.