

DOI:10.16644/j.cnki.cn33-1094/tp.2021.04.003

# 卷积神经网络研究综述

李炳臻<sup>1</sup>, 刘 克<sup>2</sup>, 顾佼佼<sup>2</sup>, 姜文志<sup>2</sup>

(1. 中国人民解放军95668部队, 云南 昆明 650000; 2. 海军航空大学岸防兵学院)

**摘 要:** 回顾了卷积神经网络的发展历程, 介绍了卷积神经网络的基本运算单元。在查阅大量资料基础上, 重点介绍了有代表性的AlexNet、VGGNet、GoogLeNet、ResNet等, 对他们所用到的技术进行剖析, 归纳、总结、分析其优缺点, 并指出卷积神经网络未来的研究方向。

**关键词:** 卷积神经网络; AlexNet; VGGNet; GoogLeNet; ResNet

中图分类号: TP399

文献标识码: A

文章编号: 1006-8228(2021)04-08-05

## Review of the researches on convolutional neural networks

Li Bingzhen<sup>1</sup>, Liu Ke<sup>2</sup>, Gu Jiaojiao<sup>2</sup>, Jiang Wenzhi<sup>2</sup>

(1. 95668 unit of PLA, Kunming, Yunnan 650000, China; 2. Naval Aviation University, Coast Guard Academy)

**Absrtact:** This paper reviews the development of convolutional neural networks, and introduces the basic operation unit of convolutional neural networks. On the basis of consulting a large amount of information, this paper focuses on the representative convolutional neural networks such as AlexNet, VGGNet, GoogLeNet and ResNet etc., analyzes the technologies they used, summarizes and analyzes their advantages and disadvantages, and points out the future research direction of convolutional neural networks.

**Key words:** convolutional neural networks; AlexNet; VGGNet; GoogLeNet; ResNet

## 0 引言

卷积神经网络(Convolutional Neural Networks, CNN)是一类包含卷积计算并且含有深层次结构的深度前馈神经网络, 是深度学习的代表算法之一<sup>[1-2]</sup>, 21世纪后, 随着深度学习理论的提出和数值计算设备的改进, 卷积神经网络得到了快速发展。较之于传统方法, 卷积神经网络的优点在于可自动提取目标特征, 发现样本集中特征规律, 解决了手动提取特征效率低下、分类准确率低的不足, 因此卷积神经网络被广泛应用于图像分类、目标识别、自然语言处理等领域<sup>[3]</sup>, 取得了瞩目的成就。

## 1 卷积神经网络的发展历程

卷积神经网络最早可以追溯到1943年, 心理学家Warren和数理逻辑学家Walter在文献[4]第一次提出

神经元的概念, 通过一个简单的数学模型将神经反应简化为信号输入、求和、线性激活及输出, 具有开创性意义; 1958年神经学家Frank通过机器模拟了人类的感知能力<sup>[5]</sup>, 这就是最初的“感知机”, 同时他在当时的计算机IBM704上完成了感知机的仿真, 能够对三角形和四边形进行分类, 这是神经元概念提出后第一次成功的实验, 验证了神经元概念的可行性。以上是神经元发展的第一阶段。第一代神经网络由于结构简单, 仅能解决线性问题。此外, 认知的限制也使得神经网络的研究止步于此。

第二代卷积神经网络出现于1985年, Geoffrey Hinton在神经网络中使用多个隐含层进行权重变换, 同时提出了误差反向传播算法(Backpropagation algorithm, BP算法)<sup>[6]</sup>, 求解各隐含层的网络参数, 优点是理论基础牢固、通用性好, 不足之处在于网络收敛速度慢、容易出

收稿日期: 2020-11-19

作者简介: 李炳臻(1996-), 男, 山西运城人, 硕士, 主要研究方向: 人工智能深度学习技术。

现局部极小的问题;1988年,Wei Zhang 提出平移不变人工神经网络<sup>[7]</sup>(SIANN),将其应用在医学图像检测领域;1989年,LeCun 构建了应用于计算机视觉问题的卷积神经网络,也就是 LeNet 的早期版本,包含两个卷积层、两个全连接层共计 6 万多个参数,结构上与现代的卷积神经网络模型结构相似,而且开创性的提出了“卷积”这一概念,卷积神经网络因此得名。1998 年 LeCun 构建了更加完备的卷积神经网络 LeNet-5<sup>[8]</sup>并将其应用于手写字识别,在原有 LeNet 的基础上加入了池化层,模型在 Mnist 上的识别准确率达到 98% 以上,但由于当时不具备大规模计算能力的硬件条件,因此卷积神经网络的发展并没有引起足够的重视。

第三代卷积神经网络兴起于 2006 年,统称为深度学习,分为两个阶段,2006 至 2012 年为快速发展期,2012 至今为爆发期,训练数据量越大,卷积神经网络准确率越高,同时随着具备大规模计算能力 GPU 的应用,模型的训练时间大大缩短,深度卷积神经网络的发展是必然的趋势。2006 年 Hintont 提出了包含多个隐含层的深度置信网络<sup>[9]</sup>(Deep Belief Network, DBN),取得了十分好的训练效果,DBN 的成功实验拉开了卷积神经网络百花齐放的序幕:自 2012 年 AlexNet<sup>[10]</sup>取得 ImageNet 视觉挑战赛的冠军,几乎每年都有新的卷积神经网络产生,诸如 ZFNet<sup>[11]</sup>、VGGNet<sup>[12]</sup>、GoogLeNet<sup>[13]</sup>、ResNet<sup>[14]</sup>以及 DPRSNet<sup>[15]</sup>等等,都取得了很好的效果。

## 2 卷积神经网络的基本运算单元

卷积神经网络中隐含层低层中的卷积层与池化层交替连接,构成了卷积神经网络的核心模块,高层由全连接层构成。

### 2.1 卷积层

卷积层用于提取输入的特征信息,由若干卷积单元组成,每个卷积单元的参数都是通过反向传播算法优化得到的,通过感受野(filter)对输入图片进行有规律的运动,并与所对应的区域做卷积运算提取特征;低层卷积只能提取到低级特征,如边缘、线条等,高层卷积可以提取更深层的特征。

卷积层参数包括 filter 大小、步长(stride)和边界填充(pad),三者共同决定了卷积层输出特征图的尺寸大小;感受野大小小于输入图片尺寸,感受野越大,可提取的特征越复杂;stride 定义了 filter 扫过相邻区域时的位置距离;pad 是在特征图周围进行填充避免输出特征丢失过多边缘信息的方法,pad 值代表填充层数。

卷积层的运算过程如图 1,输入为一个  $5 \times 5 \times 3$  的图像, pad=1, 边界填充后,尺寸为  $7 \times 7 \times 3$ , 使用  $3 \times 3 \times 3$  感受野进行卷积运算,整个运算过程可以看成卷积核输入层上滑动计算,每次滑动的距离为 2,从原图左上角开始从左至右滑动进行卷积运算,这一行处理结束后,移动到下一行,继续向右滑动,直至覆盖所有区域,共得到 9 个运算结果,输出  $3 \times 3$  的矩阵,称为特征图。表达式如下:

$$f(x) = \sum_{ij} \theta_{ij} \times x_{ij} + b \quad (1)$$

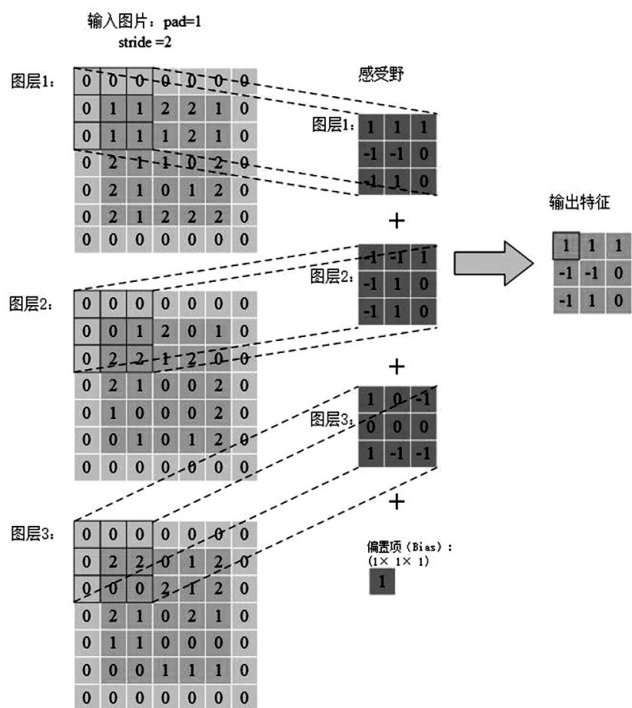


图 1 卷积核运算过程示意图

在公式(1)中, $f(x)$ 代表输出特征, $\theta_{ij}$ 代表第  $i$  行  $j$  列的卷积核元素大小, $x_{ij}$ 代表第  $i$  行  $j$  列元素大小, $b$  为偏差。卷积层具有局部感受野、权值共享的特点,可以减少网络中的参数。

### 2.2 激活函数层

卷积运算提取到的图像特征是线性的,但真正的样本往往是非线性的,为此引入非线性函数来解决。激活函数,使得每个像素点可以用 0 到 1 的任何数值来代表,模拟更为细微的变化。激活函数一般具有非线性、连续可微、单调性等特性。

比较常用的激活函数有 Sigmoid 函数、tanh 函数,以及 ReLU 函数(Rectified Linear Unit,即线性整流函数)。Sigmoid 函数的定义式是  $f(x) = \frac{1}{1 + e^{-x}}$ , tanh 函数的

定义式为  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , ReLU 函数的定义式为  $f(x) = \max(0, x)$ , 曲线如图2。

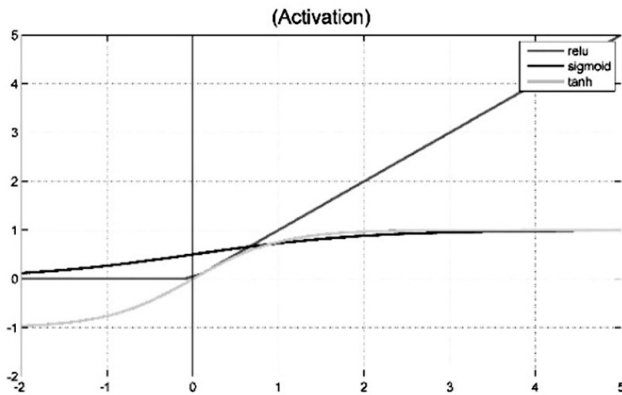


图2 三种激活函数图像

sigmoid函数单调易求导,但输出不以0为中心,且饱和时梯度非常小,网络权值得不到有效的更新;tanh函数本质上为  $2 \times \text{sigmoid}(2x) - 1$ ,仍存在饱和时梯度过小的问题;ReLU函数具有单侧抑制、稀疏激活性的特点,即在  $x > 0$  时,梯度恒为1,无梯度耗散问题; $x < 0$  时,输出为0,训练结束后神经元为0越多,网络越稀疏,提取出的特征越具有代表性,能够缓解过拟合的问题。

缺点在于强制的稀疏处理会导致模型无法学习到有效的特征。最常用是ReLU函数,但要注意学习率的设置,防止过多神经元坏死。

### 2.3 池化层

池化层的作用为压缩特征图,提取主要特征,简化网络计算的复杂度。池化方式一般有两种,均值池化与最大池化。示意如图3。

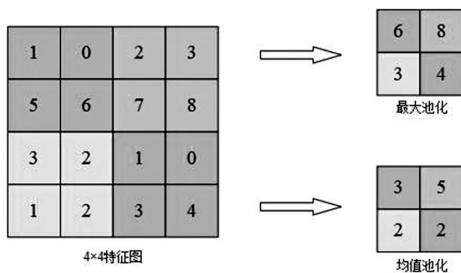


图3 池化操作示意图

图3中,采用一个  $2 \times 2$  的 filter, stride 为2, padding 为0。最大池化即在  $2 \times 2$  的区域中寻找最大值;均值池化则是求每一个  $2 \times 2$  的区域中的平均值,得到主要特征。一般最常用的 filter 取值为2, stride 为2,池化操作将特征图缩小,有可能影响网络的准确度,但可以

通过增加网络深度来弥补。

### 2.4 全连接层

全连接层位于卷积神经网络的最后,给出最后的分类结果,在全连接层中,特征图会失去空间结构,展开为特征向量,并把由前面层级所提取到的特征进行非线性组合得到输出可用公式(2)表示。

$$f(x) = W \times x + b \quad (2)$$

公式(2)中,  $x$  为全连接层的输入,  $W$  为权重系数,  $b$  为偏置。全连接层连接所有特征输出至输出层,对于图像分类问题,输出层使用逻辑函数或归一化指数函数 (softmax function) 输出分类标签<sup>[16]</sup>。在识别 (object detection) 问题中,输出层输出为物体的中心坐标、大小和分类<sup>[16]</sup>。在语义分割中,则直接输出每个像素的分类结果<sup>[16]</sup>。

## 3 几种性能优异的卷积神经网络

### 3.1 AlexNet

AlexNet 赢得了2012 ILSVRC(ImageNet 大规模视觉识别挑战赛)的冠军。2012年是CNN首次实现Top 5误差率15.4%的一年(Top 5误差率是指给定一张图像,其标签不在模型认为最有可能的5个结果中的概率),第二名使用传统识别方法得到的误差率为26.2%。卷积神经网络在这次比赛的表现震惊了整个计算机视觉界,奠定了卷积神经网络在计算机视觉领域的绝对地位。

AlexNet 包含6亿3千万个连接,6000万个参数和65万个神经元,网络结构如图4。

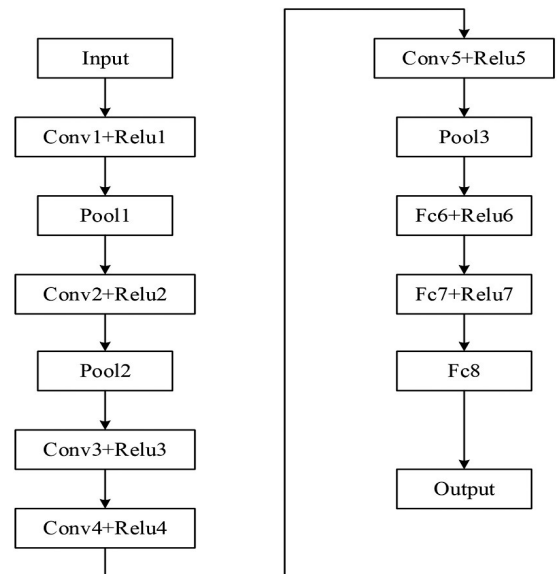


图4 AlexNet 网络结构简图



AlexNet的成功,除了深层次的网络结构,还有以下几点:首先,采用ReLU作为激活函数,避免了梯度耗散问题,提高了网络训练的速度;其次,通过平移、翻转等扩充训练集,避免产生过拟合;最后提出并采用了LRN(Local Response Normalization,局部响应归一化处理),利用临近的数据做归一化处理技术,提高深度学习训练时的准确度;除此之外,AlexNet使用GPU处理训练时所产生大量的矩阵运算,提升了网络的训练效率。

### 3.2 VGGNet

VGGNet是牛津大学与GoogleDeepMind公司的研究员一起合作开发的卷积神经网络,2014年取得了ILSVRC比赛分类项目的亚军和识别项目的冠军。VGGNet探索了网络深度与其性能的关系,通过构筑16~19层深的卷积神经网络,Top 5误差率为7.5%,在整个卷积神经网络中,全部采用 $3 \times 3$ 的卷积核与 $2 \times 2$ 的池化核,网络结构如图5。

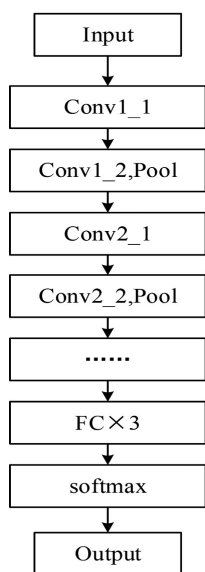


图5 VGGNet的网络结构示意图

VGGNet包含很多级别的网络,深度从11层到19层不等,最常用的是VGG-16和VGG-19。VGGNet把网络分成了5段,每段都把多个 $3 \times 3$ 的网络串联在一起,每段卷积后接一个最大池化层,最后是3个全连接层和一个softmax层。

VGGNet有两个创新点。①通过网络分段增加网络深度、采用多层小卷积代替一层大卷积,2个 $3 \times 3$ 的卷积核相当于 $5 \times 5$ 的感受野,三个相当于 $7 \times 7$ 的感受野。优势在于:首先包含三个ReLU层增加了非线性操作,对特征的学习能力更强;其次减少了参数,

使用 $3 \times 3$ 的3个卷积层需要 $3 \times (3 \times 3 \times n) = 27 \times n$ 个参数,使用 $7 \times 7$ 的一个卷积层需要 $7 \times 7 \times n = 49 \times n$ 个参数。②在训练过程中采用了多尺度和交替训练的方式,同时对一些层进行预训练,使得VGGNet能够在较少的周期内收敛,减轻了神经网络训练时间过长的问題。不足之处在于使用三个全连接层,参数过多导致内存占用过大,耗费过多的计算资源。

VGGNet是最重要的神经网络之一,它强调了卷积网络的深度的增加对于性能的提升有着重要的意义。

### 3.3 GoogLeNet

GoogLeNet是由谷歌研究院提出的卷积神经网络,获得2014年的ILSVRC比赛分类任务的冠军,Top 5误差率仅为6.656%。GoogLeNet的网络共有22层,但参数仅有700万个,比之前的网络模型都少很多。一般来说,提升网络性能最直接的办法就是增加网络深度,随之增加的还有网络中的参数,但过量的参数容易产生过拟合,也会增大计算量。GoogLeNet采用稀疏连接解决这种问题,为此提出了inception的结构,如图6所示。

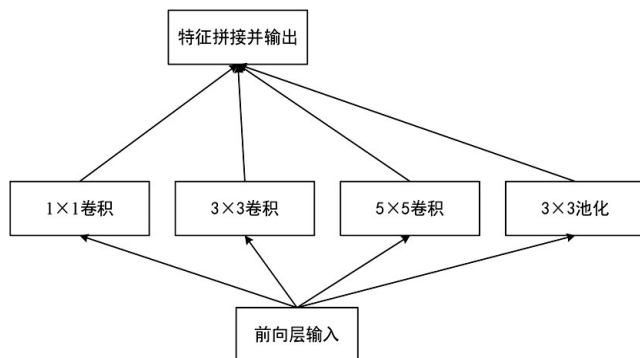


图6 GoogLeNet中inception结构示意图

在inception结构中,同时采用 $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$ 卷积核是为了将卷积后特征图保持一致,便于融合, stride=1, padding 分别为0、1、2,卷积后就可得到相同维度的特征,最后进行拼接,将不同尺度的特征进行融合,使得网络可更好的提取特征。

在整个网络中,越靠后提取到的特征也越抽象,每个特征所对应的感受野也随之增大,因此随着层数的增加, $3 \times 3$ 、 $5 \times 5$ 卷积核的比例也要随之增加,这样也会带来巨大的参数计算,为此GoogLeNet也有过诸多改进版本,GoogLeNet Inception V2、V3以及V4,通过增加Batch Normalization、在卷积之前采用 $1 \times 1$ 卷积降低纬度、将 $n \times n$ 的卷积核替换为 $1 \times n$ 和 $n \times 1$ 等方法降低网络参数,提升网络性能。

### 3.4 ResNet

ResNet 于 2015 年被提出, 获得了 ILSVRC 比赛的冠军, ResNet 的网络结构有 152 层, 但 Top 5 错误率仅为 3.57%, 之前的网络都很少有超过 25 层的, 这是因为随着神经网络深度的增加, 模型准确率会先上升后达到饱和, 持续增加深度时, 准确率会下降; 因为随着层数的增多, 会出现梯度爆炸或衰减现象, 梯度会随着连乘变得不稳定, 数值会特别大或者特别小; 因此网络性能会变得越来越差。ResNet 通过在网络结构中引入残差网络来解决此类问题, 残差网络结构示意图如图 7 所示。

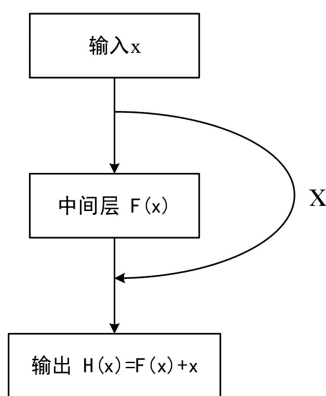


图 7 残差网络结构示意图

很明显, 残差网络是跳跃结构, 残差项原本是带权值的, 但 ResNet 用恒等映射代替了它。在图 7 中, 输入为  $x$ , 期望输出为  $H(x)$ , 通过捷径连接的方式将  $x$  传到输出作为初始结果, 输出为  $H(x) = F(x) + x$ , 当  $F(x) = 0$  时,  $H(x) = x$ 。于是, ResNet 相当于将学习目标改变为目标值  $H(x)$  和  $x$  的差值, 也就是所谓的残差  $F(x) = H(x) - x$ , 因此, 后面的训练目标就是要将残差结果逼近于 0。

ResNet 通过提出残差学习, 将残差网络作为卷积神经网络的基本结构, 通过恒等映射, 来解决因网络模型层数过多导致的梯度爆炸或衰减问题, 可以最大程度地加深网, 并得到非常好的分类效果。

### 4 结束语

本文对卷积神经网络的发展进行了梳理和分析, 归纳、总结了卷积神经网络中基本的运算单元, 重点剖析了几个性能优异应用广泛的卷积神经网络, 卷积神经网络目前已得到了广泛的应用, 但仍然存在大量值得研究的热点方向:

(1) 如何设计结构可以使得模型具备更高层次的

智能化, 从而解决更加复杂的问题;

(2) 随着卷积神经网络层数的不断加深, 卷积神经网络的训练, 势必会消耗大量能耗、占用大量内存, 应用在生产中如何降低能耗, 降低生产成本;

(3) 手机的大力普及对于卷积神经网络的实时性需求增加, 如何设计轻量级、小巧的卷积神经网络, 同时保证实时性与模型性能也是值得研究的问题。

### 参考文献(References):

- [1] Goodfellow, I., Bengio, Y., Courville, A.. Deep learning (Vol. 1). Cambridge: MIT press, 2016: 326–366
- [2] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang, G. and Cai, J., 2015. Recent advances in convolutional neural networks. arXiv preprint arXiv:1512.07108
- [3] Sun B, Ju Q Q, Sang Q B. Image dehazing algorithm based on FC-DenseNet and WGAN[J]. Journal of Frontiers of Computer Science and Technology, 2020.14(8): 1380–1388
- [4] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943.5(4): 115–133
- [5] Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain[J]. Psychological Review, 65(6): 386–408
- [6] Rumelhart D E, Hinton G E, Williams R J. Learning Representations by Back Propagating Errors[J]. Nature, 1986.323(6088): 533–536
- [7] Zhang, W., 1988. Shift-invariant pattern recognition neural network and its optical architecture. In Proceedings of annual conference of the Japan Society of Applied Physics.
- [8] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278–2324
- [9] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief net[J]. Neural Comput, 2006.18(7): 1527–1554
- [10] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems, 2012: 1097–1105
- [11] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//European conference on computer vision. Springer, Cham, 2014: 818–833

(下转第 17 页)



图8 Prometheus+Grafana 监控示意图



图9 Email告警示意图

## 4 结束语

本文重点介绍了DevOps概念、优势以及实现的技术路线,最后结合工程数据中心项目描述DevOps落地应用情况。可以看到,DevOps不仅极大的提高了需求管理的准确性,减少了开发和运维的沟通成本,提升了软件系统的健壮性,实现了研发过程管理提质

增效,最终从多个方面完善了公司的研发体系。

### 参考文献(References):

- [1] 荣国平,白晓颖,岳涛.面向DevOps的软件工程新技术专题前言[J].软件学报,2019.10.
- [2] 张文林.持续交付及其在大型项目中的应用[J].软件导刊,2017.10.
- [3] 王敏,贾百强,李翔,李小军.基于JIRA的全生命周期软件开发管理平台[J].上海船舶运输科学研究所学报,2017.2.
- [4] 蔡永健,路云菲,邹远祥,唐海清.基于Jenkins和Docker容器技术在数字化电站项目自动化部署的研究和应用[J].计算机时代,2020.2.
- [5] 付金莹,蒋云,王金伦,申宗杰.基于静态分析的应用代码质量管理体系研究及应用[J].中国新通信,2013.21.
- [6] 石宇婷.基于Zabbix平台的私有云监控系统设计与实现[D].东华大学,2017.
- [7] 马永,吴跃,何孝国,程航.基于Prometheus的基础软硬件全链路监控设计和实现[J].电子技术与软件工程,2020.1.

(上接第12页)

- [12] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large Scale Image Recognition[J]. Computer ence, 2014.
- [13] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition, 2015.
- [14] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of

the IEEE conference on computer vision and pattern recognition,2017:4700-4708

- [15] Yamada Y, Iwamura M, Kise K. Deep pyramidal residual networks with separated stochastic depth[J]. arXiv preprint arXiv:1612.01230, 2016.
- [16] 尚挺,费建超,周游,方虎生,朱经纬.基于深度卷积神经网络的行人检测[J].计算机工程与应用,2016.52(13):162-166