计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(4):247−252 [doi: 10.15888/j.cnki.csa.007535]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

面向复杂验证码识别任务的轻量神经网络设计①

李　昊, 程　辉

(江汉大学 数学与计算机科学学院, 武汉 430056)
通讯作者: 程　辉, E-mail: 497175101@qq.com

摘　要: 深层神经网络拥有更强特征表达能力的同时, 也带来了优化难、训练成本高及梯度弥散等问题; 参数数量的激增则导致模型过于臃肿, 不利于其在移动端及工业控制设备等算力弱、存储小的平台上的部署. 针对这些问题, 构建了一种融合空洞卷积和多尺度稀疏结构的轻量神经网络对图像进行特征提取, 实现对带有彩色图形噪声且字符扭曲粘连严重的验证码图像的端到端识别. 将包含 100 万张验证码图像的数据集按 98:1:1 的比例划分为训练集、验证集和测试集, 逐批参与训练. 实验结果表明, 该网络在大大减少参数数量的同时, 具有测试集上 98.9% 的识别成功率.

关键词: 轻量化; 卷积神经网络; 多尺度稀疏网络结构; 空洞卷积

引用格式: 李昊,程辉.面向复杂验证码识别任务的轻量神经网络设计.计算机系统应用,2021,30(4):247–252. http://www.c-s-a.org.cn/1003-3254/7535.html

Lightweight Neural Network Design for Complex Verification Code Recognition Task

LI Hao, CHENG Hui

(School of Mathematics and Computer Science, Jianghan University, Wuhan 430056, China)

Abstract: The deep neural network can better express features but results in difficult optimization, high training cost, and vanishing gradient. The surge in quantity of parameters leads to a too bloated model to be deployed on the platform with weak computing power and small storage, such as mobile terminal and industrial control equipment. Aiming at these problems, we construct a lightweight neural network combining atrous convolutions and multi-scale sparse structures to extract the features of images, and realize the end-to-end recognition for the captcha images with color pattern noise and seriously touched and distorted characters. The dataset containing one million images was divided into training sets, validation sets, and test sets in the ratio of 98:1:1 and trained in batches. Consequently, the lightweight neural network has a recognition rate of 98.9% on test sets with much fewer parameters.

Key words: lightweight; Convolutional Neural Network (CNN); multi-scale sparse structure; atrous convolutions

验证码是一种被广泛应用于互联网安全领域, 可有效防止自动化脚本对服务器进行攻击、区分人类和计算机的图灵测试技术[1]. 较低的分辨率及背景带干扰等特点, 使其对传统的光学字符识别技术 (Optical Character Recognition, OCR) 具有一定的抵抗能力, 可有效避免金融诈骗、刷单式营销等恶性事件的发生, 因此备受金融证券系统和电商网站的青睐, 被广泛部署于各项服务中, 为建设安全的互联网环境做出了巨大的贡献. 近年来, 计算机视觉技术 (Computer Vision, CV) 飞速发展, 传统的验证码变得不再安全. 由此引发了验证码生成技术的更新迭代, 出现了带有彩色图形噪声、字符扭曲粘连严重的新型验证码. 对于这种情况, 旋转匹

① 基金项目: 国家自然科学基金 (61472148, 61701194); 湖北省教育厅科研计划 (B2018254)
Foundation item: National Natural Science Foundation of China (61472148, 61701194); Scientific Research Program of Education Bureau of Hubei Province (B2018254)
收稿时间: 2020-01-09; 修改时间: 2020-02-08; 采用时间: 2020-02-24; csa 在线出版时间: 2021-03-30

配、支持向量机 (Support Vector Machine, SVM)[2,3] 等传统算法变得不再适用.

得益于脑科学和神经科学的发展, 有学者提出通过构造人工神经网络 (Artificial Neural Network, ANN), 即把众多具有信息处理功能的“神经元”按照一定的方式进行连接, 组成网络反复训练, 来模仿人脑的学习过程, 进而达成“机器以人的方式学习图片特征, 识别验证码”的效果. 1998 年, LeCun 等人[4] 在已有的工作基础上设计了如图 1 所示的名为 LeNet-5 的卷积神经网络 (Convolutional Neural Network, CNN) 并在手写数字的识别任务上取得成功. 这之后, 陆续有研究者设计出更深层的网络. 其中, 文献 [5] 设计的 AlexNet 首先将修正线性单元 ReLU 作为网络各层的激活函数, 大大缩短了模型收敛所需的时间; 文献 [6] 提出的 VGG16 和 VGG19 则在加深网络层数的同时使用小卷积核代替大卷积核, 不仅减少了卷积操作带来的计算量, 还提升了模型提取图像特征的精度; 文献 [7] 设计的 GoogLeNet 提出使用多尺度稀疏结构, 保持网络结构稀疏性的同时, 充分利用了密集矩阵的高计算性能. 实验证明, 将它们用于验证码识别都可取得不错的成绩. 卷积神经网络深度的增加带来更强的图像特征表达能力的同时, 优化难、训练成本高等问题也接踵而至, 参数数量的激增, 又令模型变得十分“臃肿”, 限制了其在移动端及工业控制设备等算力弱、存储小的平台上的部署. 如何在减少参数数量的同时获得更强的特征表示能力和更高的识别成功率, 便是一个十分重要的问题.

基于上述分析, 本文设计了一种融合了空洞卷积和多尺度稀疏结构的轻量神经网络用于复杂验证码图像的特征提取和识别, 计算网络输出的 Sigmoid 交叉熵作为模型损失率, 采用 RAdam 算法做优化, 其特点及计算过程在第 1.3 节中介绍.

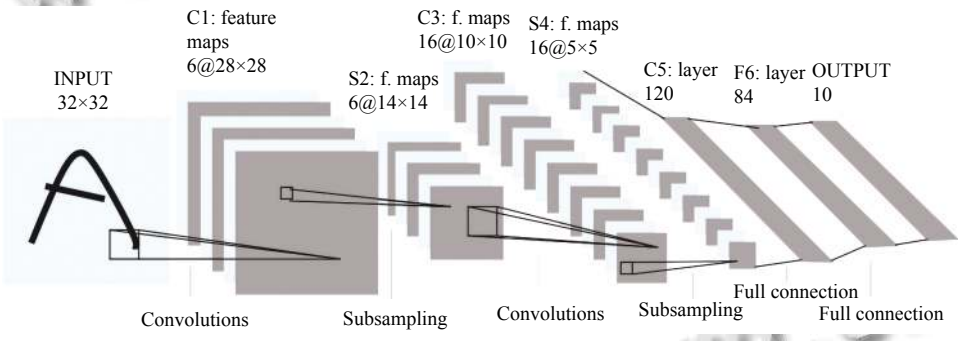


图 1 LeNet-5 的结构

1 设计方案

1.1 图像预处理模块

对图像进行适当的预处理, 可有效消除图像中无关识别的信息[8], 降低维度, 既增强了特征因素的可检测性, 也有助于神经网络设计的轻量化. 本次实验中对原始样本集的预处理由统一尺寸及灰度转换两部分组成. 其中, 灰度转换将三通道的验证码图像转换为单通道[9], 这使得参与计算的数据量大大减少, 既有利于图像的进一步处理, 也降低了神经网络的负载, 使其变得更轻. 其转换方法可用式 (1) 表示:

$$Gray \leftarrow (19\,595R + 38\,469G + 7472B) \gg 16 \quad (1)$$

1.2 轻量化主干特征提取模块

轻量神经网络的设计要遵循几个原则: 尽量完整的网络结构、不宜过多的层数以及尽可能少的内存访问成本 (Memory Access Cost, MAC)[10]. 基于此, 本文设计了如图 2 所示的名为 LNet 的轻量神经网络, 其参数信息如表 1 所示. 其中, 卷积操作使用如图 3(a) 所示的大小为 3×3, 扩张率为 2 的空洞卷积核 (atrous convolutions), 相比于如图 3(b) 所示的传统卷积核, 可在不增加参数的前提下获得更大的感受野 (Receptive Field, RF)[11]; 使用 1×1 卷积核在不损害模型表达能力的前提下大大减少参数量和计算量[12]. 使用 Leaky ReLU[13] 作为各层的激活函数来缓解梯度消失问题, 不使用修正效果更好的 PReLU 函数的原因在于 PReLU 会引入新的超参, 加重模型的训练负担. Leaky ReLU 的修正方法可用式 (2) 表示 (a 为常数).

$$f(x) \leftarrow \max(ax, x) \quad (2)$$

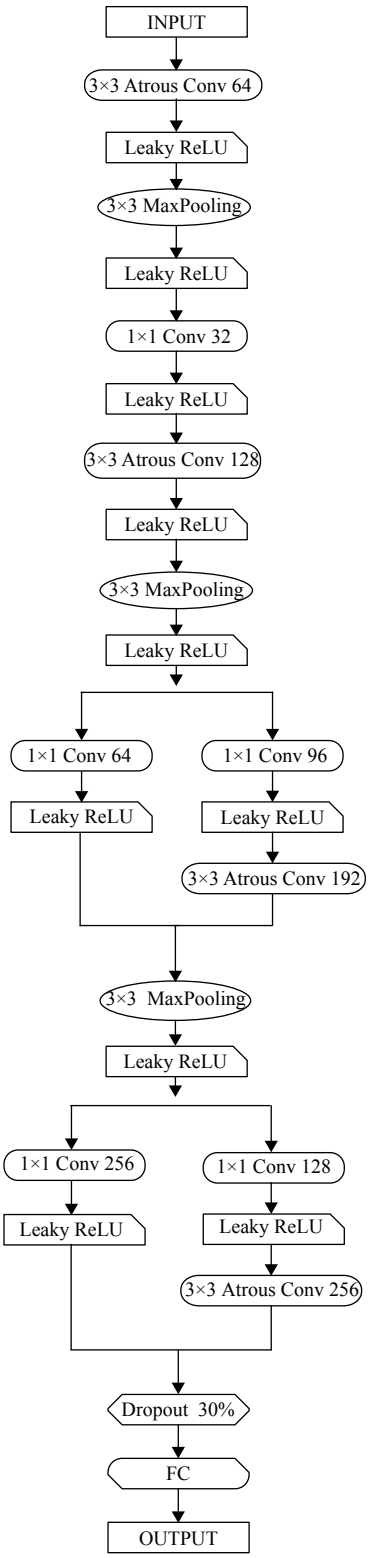


图 2 轻量神经网络

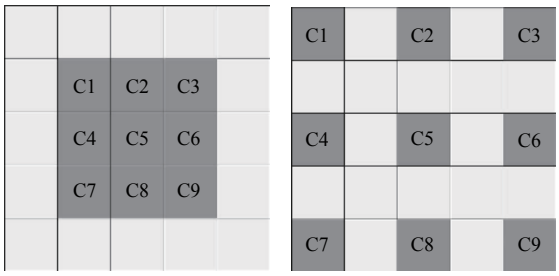
1.3 损失函数及优化方法

交叉熵在信息论中可用于度量两个概率分布的差异性. 适用于验证码识别任务的 Sigmoid 交叉熵作为损失函数可用式 (3) 表示.

$$C \leftarrow -1/n \sum [y \ln a + (1-y)\ln(1-a)] \tag{3}$$

表 1 LNet 参数信息

网络层	核尺寸	输出
Input	/	224×224×1
Convolutional	3×3>64(Atrous)	224×224×64
MaxPooling	3×3	112×112×64
Convolutional	1×1>32	112×112×32
Convolutional	3×3>128(Atrous)	112×112×128
MaxPooling	3×3	56×56×128
Convolutional	1×1>64 1×1>96 3×3>192(Atrous)	56×56×256
MaxPooling	3×3	28×28×256
Convolutional	1×1>256 1×1>128 3×3>256(Atrous)	28×28×512
Dropout, 32%	/	/
FC	/	1024
Output	/	372



(a) 普通卷积核 (b) 空洞卷积核

图 3 普通卷积核和空洞卷积核

RAdam 由 Liu 等人[14] 提出, 兼有 Adam 优化器[15–17] 和随机梯度下降法 (Stochastic Gradient Descent, SGD)[18] 两者的优点, 可在获得较快收敛速度的同时有效避免掉入局部最优解的陷阱. 根据方差分散度动态地打开或关闭自适应学习率, 在大数据集上表现良好且内存需求不高, 其计算过程可用式 (4) ~ 式 (12) 表示:

$$g_t \leftarrow \Delta_\theta f_t(\theta_{t-1}) \tag{4}$$

$$m_t \leftarrow \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{5}$$

$$v_t \leftarrow \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \tag{6}$$

$$\hat{m}_t \leftarrow m_t / \left(1-\beta_1^t\right) \tag{7}$$

$$\rho_t \leftarrow \rho_\infty - 2t\beta_2^t / \left(1-\beta_2^t\right) \tag{8}$$

$$\text{If } (\rho_\infty > 4): \quad \hat{v}_t \leftarrow \sqrt{v_t / \left(1-\beta_2^t\right)} \tag{9}$$

$$r \leftarrow \sqrt{\frac{(\rho_t - 4)(\rho_t - 4)\rho_\infty}{(\rho_\infty - 4)(\rho_\infty - 2)\rho_t}} \tag{10}$$

$$\theta_t \leftarrow \theta_{t-1} - \partial_t r_t \hat{m}_t / \hat{v}_t \quad (11)$$

$$\text{Else}: \quad \theta_t \leftarrow \theta_{t-1} - \partial_t \hat{m}_t \quad (12)$$

其中, 式 (4) 计算第 i 步时的梯度; 式 (5) 和式 (6) 分别计算移动量的一阶矩估计和二阶矩估计; 式 (7) 对一阶矩估计的值做偏差修正; 式 (8) 计算 SMA (Simple Moving Average) 的最大长度; 式 (9) 对二阶矩估计的值做偏差修正; 式 (10) 获得方差的修正范围.

2 数据集与实验环境

本文借助爬虫工具收集了 100 万张如图 4 所示的带有彩色图形噪声、字符扭曲粘连的验证码图像建立主数据库 IDB1, 其中每张图像的原始大小为 180×92; 分别采集来自大连理工大学综合教务系统和微信开发者平台的 6 万张验证码图像建立数据库 IDB2 和 IDB3, 每张图像的原始大小为 168×84. 经预处理后的样本按照 98:1:1 的比例划分为训练集 (train set)、验证集 (validation set) 和测试集 (test set). 其中训练集用于训练模型, 在梯度下降 (gradient descent) 的过程中确定网络各层的权重和偏置等参数; 验证集用于学习率、深度等超参数 (hyperparameter) 的优化[19]; 测试集则用于对模型性能的最终评估.

图 4 超分辨率重建后的验证码图像示例

本文模型构建及算法设计基于开源框架 PyTorch 实现, 实验中模型训练的初始学习率 (learning rate) 为 0.003. 采用随机遍历算法确保训练集中的每一张图像都可被学习到, 每次处理 256 张图像, 最大迭代轮数 (epochs) 为 500. 开展实验的软硬件环境如表 2 所示.

表 2 实验环境的配置参数

参数	值	参数	值
操作系统	Ubuntu 16.04	CUDA版本	9.2
CPU	i7 9700 K 3.60 GHz	CUDNN版本	7.6
GPU	Tesla V100/16 GB	PyTorch版本	0.4.0
RAM	32 GB/DDR4	Python版本	3.6

3 训练集和测试集上的拟合情况

LNet 在训练集和测试集上 accuracy 和 loss 的拟合曲线分别如图 5 和图 6 所示. 可知拟合状况良好.

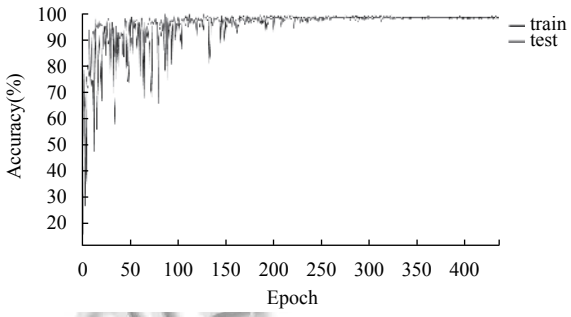


图 5 训练集和测试集上的 accuracy 曲线

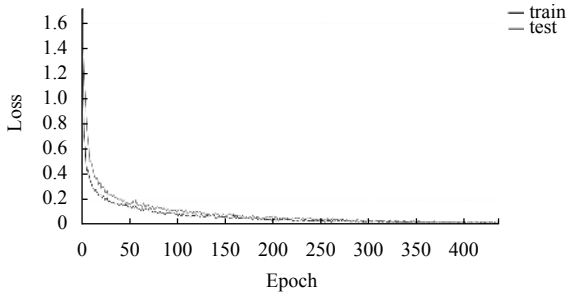


图 6 训练集和测试集上的 loss 曲线

4 对比实验结果及分析

4.1 图像预处理模块

为验证本文设计的复杂验证码识别模型的优越性, 设计了本文模型与基于 VGG16 设计的模型 A、基于 ResNet-18 设计的模型 B、基于 GoogLeNet 设计的模型 C 和基于生成对抗网络 (Generative Adversarial Networks, GAN) 设计的模型 D 的对比实验. 实验在相同的实验环境下展开, 使用相同的训练集训练模型. 5 种模型在 IDB1、IDB2 和 IDB3 测试集上的实验结果如表 3 和表 4 所示.

综合表 3 和表 4 可知, 本文设计的验证码识别模型除了在 IDB2 上的识别成功率小于基于 GAN 设计的方法外, 综合性能较其余 4 种模型均有着一定优势.

表 3 5 种模型识别成功率 (%)

模型	IDB1	IDB2	IDB3
A (VGG16)	96.4	89.2	93.6
B (ResNet-18)	89.3	82.5	88.7
C (GoogLeNet)	92.6	88.4	94.5
D (GAN)	97.2	**91.8**	96.1
E (Ours LNet)	**98.9**	90.9	**96.4**

表 4 5 种模型的识别平均时间 (单位: s)

模型	IDB1	IDB2	IDB3
A (VGG16)	2.03	1.99	2.01
B (ResNet-18)	1.33	1.25	1.41
C (GoogLeNet)	1.62	1.39	1.47
D (GAN)	2.62	2.12	2.36
E (Ours LNet)	**1.08**	**0.86**	**1.01**

4.2 消融实验

为说明本文在设计轻量神经网络时采用的策略对模型性能的影响, 设计了空洞卷积、1×1 卷积核和多尺度稀疏结构在 IDB1 主数据库上的消融实验: 针对空洞卷积, 将 LNet 中用于验证码图像特征提取的卷积层改用尺寸为 3×3 的普通卷积核, 验证增大感受野对模型性能的影响; 针对 1×1 卷积, 删除所有用于降维的 1×1 卷积层, 查看其对参数数量的影响; 针对多尺度稀疏网络结构, 将其更换为如图 7(b) 所示的传统结构. 采用参数数量和识别率两项指标评估模型性能, 实验结果如表 5 所示. 参数数量的计算方法[20,21] 可用式 (13) 表示.

$$Parameters \leftarrow \sum_{i}^{D} K_{il}^2 \cdot C_{iI} \cdot C_{iO} \tag{13}$$

其中, D 是神经网络的层数, i 表示第 i 个卷积层, k_{il} 表示第 i 个卷积层卷积核的边长, C_{iI} 表示输入第 i 个卷积层的通道数, C_{iO} 表示第 i 个卷积层输出的通道数. 模型的参数数量直接影响算法的时间复杂度和空间复杂度, 时间复杂度则决定了模型的训练和预测时间.

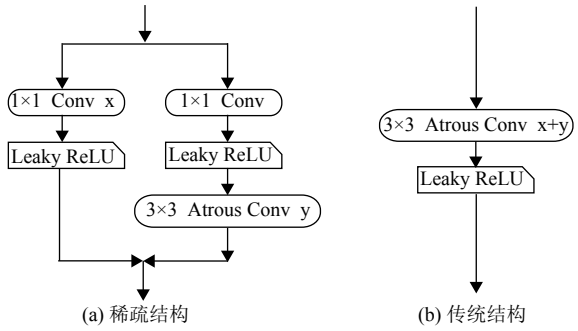


图 7 多尺度稀疏结构和传统结构

表 5 消融实验: 设计策略对模型性能的影响

空洞卷积	1×1卷积	稀疏结构	参数量(M)	识别率(%)
×	√	√	5.74	96.4
√	×	√	9.59	88.6
√	√	×	15.49	92.8
√	√	√	**5.74**	**98.7**

注: 只统计由卷积操作引入的参数数量

由表 5 可看出, 引入空洞卷积后, 网络在不增加参数量的前提下提高了 2.3% 的性能; 1×1 卷积的使用不仅提高了模型的识别率, 还将参数量减少了 40.1%; 多尺度稀疏网络结构的使用, 也对模型性能的提升大有裨益.

5 结束语

面向带彩色图形噪声、字符扭曲粘连的复杂验证码图像的识别任务, 本文融合了空洞卷积、1×1 卷积以及多尺度稀疏网络结构的优点, 设计了名为 LNet 的轻量神经网络模型. 在相同的实验环境下使用相同的训练集证明了本文设计的复杂验证码识别模型的优越性和有效性; 设计消融实验展示了不同设计策略对模型性能的影响, 证明了本文设计模型的合理性. 值得进一步讨论的还有深度可分离卷积、可变形卷积[22] 等对图像特征提取能力和模型参数数量的影响. 同时, 如何进一步地压缩模型以及如何利用现场可编程门阵列 (Field Programmable Gate Array, FPGA) 加速不同的卷积操作, 也是我们团队所关注的问题.

6 致谢

本文工作受国家自然科学基金、湖北省教育厅科研计划及国家级大学生创新创业训练计划资助, 谨在此对提供基金支持的各位老师致以崇高的敬意; 对为本文实验设计做邮件指导的浙江大学蔡登教授和为实验开展提供必要环境的中国科学院武汉水生生物研究所王莹博士致以由衷的感谢.

参考文献

1 Bursztein E, Martin M, Mitchell J. Text-based CAPTCHA strengths and weaknesses. Proceedings of the 18th ACM Conference on Computer and Communications Security. Chicago, IL, USA. 2011. 125–138. [doi: 10.1145/2046707.2046724]

2 Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293. [doi: 10.1023/A:1018628609742]

3 Avidan S. Support vector tracking. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Kauai, HI, USA. 2001. [doi: 10.1109/cvpr.2001.990474]

4 LeCun Y, Jackel L, Bottou L, *et al*. Comparison of learning

algorithms for handwritten digit recognition. International Conference on Artificial Neural Networks. Paris, France. 1995. 53–60.

5 Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Red Hook, NY, USA. 2012. 1097–1105.

6 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. https://arxiv.org/abs/1409.1556. [2014-09-04].

7 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015. 1–9. [doi: 10.1109/cvpr.2015.7298594]

8 Hu M, Yang SY. Overview of image mining research. 2010 5th International Conference on Computer Science & Education. Hefei, China. 2010. 1868–1870. [doi: 10.1109/iccse.2010.5593813]

9 Fridrich J, Goljan M, Du R. Detecting LSB steganography in color, and gray-scale images. IEEE Multimedia, 2001, 8(4): 22–28. [doi: 10.1109/93.959097]

10 Ma NN, Zhang XY, Zheng HT, *et al*. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. Proceedings of the 15th European Conference on Computer Vision (ECCV) 2018. Munich, Germany. 2018. 122–138. [doi: 10.1007/978-3-030-01264-9_8]

11 Chen LC, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848. [doi: 10.1109/tpami.2017.2699184]

12 Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. 2017. 1800–1807. [doi: 10.1109/cvpr.2017.195]

13 Xu B, Wang NY, Chen TQ, *et al*. Empirical evaluation of rectified activations in convolutional network. https://arxiv.org/abs/1505.00853. [2015-05-05].

14 Liu LY, Jiang HM, He PC, *et al*. On the variance of the adaptive learning rate and beyond. https://arxiv.org/abs/1908.03265?context=cs. [2019-08-08].

15 Stooke A, Abbeel P. Accelerated methods for deep reinforcement learning. https://arxiv.org/abs/1803.02811. [2018-03-07].

16 Wang SL, Suo SM, Ma WC, *et al*. Deep parametric continuous convolutional neural networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 2589–2597. [doi: 10.1109/CVPR.2018.00274]

17 杨观赐, 杨静, 李少波, 等. 基于 Dopout 与 ADAM 优化器的改进 CNN 算法. 华中科技大学学报 (自然科学版), 2018, 46(7): 122–127. [doi: 10.13245/j.hust.180723]

18 Ruder S. An overview of gradient descent optimization algorithms. https://arxiv.org/abs/1609.04747. [2016-09-15].

19 Guyon I. A scaling law for the validation-set training-set size ratio. AT & T Bell Laboratories, 1997: 1.

20 Bergstra J, Bengio Y. Random search for hyper-parameter optimization. Journal of Machine Learning Research, 2012, 13: 281–305.

21 唐玮, 赵保军, 龙腾. 基于轻量化网络的光学遥感图像飞机目标检测. 信号处理, 2019, 35(5): 768–774. [doi: 10.16798/j.issn.1003-0530.2019.05.005]

22 Dai JF, Qi HZ, Xiong YW, *et al*. Deformable convolutional networks. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. 2017. 764–773. [doi: 10.1109/iccv.2017.89]