

doi: 10.3969/j.issn.1674-8425(z).2020.05.021

本文引用格式: 闫河, 罗成, 李焕, 等. 基于步态能量图与 VGG 结合的步态识别方法[J]. 重庆理工大学学报(自然科学) 2020, 34(5): 166-172.

Citation format: YAN He, LUO Cheng, LI Huan, et al. Gait Recognition Method Based on Gait Energy Map Combined with VGG[J]. Journal of Chongqing University of Technology(Natural Science) 2020, 34(5): 166-172.

基于步态能量图与 VGG 结合的步态识别方法

闫 河, 罗 成, 李 焕, 李 彦

(重庆理工大学 计算机科学与工程学院, 重庆 400054)

摘 要: 基于深度学习端到端和多层特征提取的思想, 给出一种基于步态能量图和 VGG 卷积神经网络结合的步态识别方法。首先, 使用背景减除法分割出人体轮廓; 然后, 通过身体轮廓宽度变化计算出步态周期; 其次, 根据步态周期图像计算出步态能量图; 最后使用 VGG 网络对步态能量图进行特征学习及分类。实验结果表明: 所提出的方法可以准确识别行人身份, 在 CASIA-B 步态数据中平均准确率可达 92.5%, 且对视角有较好的鲁棒性, 对深度学习在步态识别领域的进一步应用有借鉴意义。

关 键 词: 步态识别; 步态能量图; 卷积神经网络

中图分类号: TP391

文献标识码: A

文章编号: 1674-8425(2020)05-0166-07

Gait Recognition Method Based on Gait Energy Map Combined with VGG

YAN He, LUO Cheng, LI Huan, LI Yan

(College of Computer Science and Engineering,
Chongqing University of Technology, Chongqing 400054, China)

Abstract: Based on the idea of end-to-end deep learning and multi-layer feature extraction, this paper adopts a gait recognition method based on gait energy map and VGG convolutional neural network. Firstly, the background contour is used to segment the human body contour, and then the gait cycle is calculated by the body contour width variation. Secondly, the gait energy map is calculated according to the gait cycle image. Finally, VGG network is used to study and classify gait energy map. The experimental results show that the proposed method can accurately identify pedestrian identity, and the average accuracy rate in the CASIA-B gait data can reach 92.5%, and it

收稿日期: 2019-02-25

基金项目: 国家自然科学基金面上项目(61173184); 重庆市自然科学基金项目(cstc2018jcyjA2328); 重庆市技术创新与应用示范项目(cstc2018jszx-cyztzx0206)

作者简介: 闫河, 男, 教授, 主要从事模式识别、目标跟踪和人工智能研究; 通讯作者: 罗成, 女, 硕士研究生, 主要从事步态识别研究, E-mail: 845570291@qq.com。

has better robustness to the viewing angle. It can be used for reference in the further application of deep learning in gait recognition.

Key words: gait recognition; gait energy image; convolutional neural network

步态识别通过人行走的方式来识别人的身份。不同于目前基于图像的个人识别方法如指纹、人脸或虹膜等,步态是唯一可在远距离非受控状态下获得而且不需要被识别人配合的生物特征,可以在较远的距离下进行检测、跟踪和识别。心理学证据表明,每个人的步态存在一定的差异,因此可用于身份识别^[1]。然而步态也有一些局限性,它可能受到服装、鞋子或环境等影响。并且,特殊的身体状况如伤病、怀孕等也可以改变一个人的步行风格。同一个人在不同条件下的步态变化降低了步态作为生物特征的识别能力,并且可能不如指纹或虹膜那样独特,但个体固有的步态特征以及随着视频监控设备在机场、车站、商场的广泛应用,步态识别在社会安全、生物识别认证和视频监控等领域逐渐发挥着越来越重要的作用。

目前,步态识别方法主要分为基于模型的步态识别方法和基于非模型的步态识别方法^[2]。基于模型的步态识别方法通常尝试对人体结构或运动方式建模,如 lee 等^[3]将人体分为七个椭圆区域,根据模板匹配的方法,分别提取每个区域的特征,然后将其组合成用于识别的步态图像特征。Bobick 等^[4]根据几何相似模型分别提取人体特征组合和身高与腿长的特征。胡荣等^[5]提出了一种基于骨架模型的步态识别方法。基于非模型的步态识别方法主要利用步态模板提取一个步态周期的轮廓特征。步态能量图^[6](GEI)是基于非模型的步态识别方法中最常见的步态模板。GEI 是通过简单地平均步态周期中的步态轮廓而生成的。其他基于非模型的步态模板表示包括计时步态图(CGI)^[7]、步态流图(GFI)^[8]和步态熵图(GEnI)^[9]。CGI 是步态帧中的时间信息由颜色映射函数编码的时间模板,CGI 通过在步态周期中组合颜色编码的步态轮廓图像来获得。GFI 是利用光流在一个步态周期中聚合灰度轮廓而生成的。GEnI 表示轮廓图像序列中像素值的随机性。但

是 Iwama 等^[10]的研究表明,在步态识别中 GEI 相比于其他的步态模板能更好地表征个体特性的综合特征,而且对噪声等有较好的鲁棒性。

近年来,深度学习以其端到端的训练方式及自主学习的特性在图像识别领域获得较好的识别效果。因此步态识别领域也逐渐引入了深度学习,用于自动提取步态特征,对行人进行分类识别。卷积神经网络(CNN)是最常用的深度学习模型之一,它是一种生物启发式的多层神经网络,通过卷积来实现特征提取,并且通过卷积的共享权值及池化,来减小网络参数的数量级,最后使用传统分类器完成分类等任务。Shiraga^[11]等设计了 GEINet,GEINet 为 8 层 CNN 网络,其中前 6 层是卷积、池化和归一化两个连续重复,再加上两个全连接层。用步态能量图作为 CNN 的输入,在 OU-ISIR 数据集上平均 rank1 准确率达到 91.6%。尽管 GEINet 达到一个较好的识别准确率,但是该网络层数较浅,步态能量图中复杂的步态特征没有提取充分,并且该网络只说明对视角有较好的鲁棒性,没有考虑到服装、携带物等因素的影响。经典网络模型 VGG^[12-14]是牛津视觉几何组开发的卷积神经网络,在 2014 年 ILSVRC 竞赛中获得第二名,相比于之前的卷积神经网络如 AlexNet 有小卷积核、小池化核、层数更深特征图更宽和全连接转卷积等特点。杨亚楠等^[15]针对传统的接收信号强度指示(RSSI)测距算法过于依赖经验或测距环境从而导致测距精度不高的问题,提出了一种基于卷积神经网络的测距算法。将 RSSI 样本数据集作为输入训练卷积神经网络,并通过该网络对实时捕获的 RSSI 信号进行特征提取,从而估计节点之间的距离。李耀龙等^[16]针对传统机场及机场内飞机目标的识别方法特征难以设计选取、泛化能力差以及难以实现两种目标同时识别的不足,将深度学习的目标识别 YOLO(You only look once)算法应用到机场及机场内飞机目标的识别

领域,实现了两种目标的同时识别。基于上述特点,本文尝试将 VGG 网络模型应用到步态识别中,探讨深卷积层和小卷积核对步态识别中特征提取和分类的影响。实验结果表明,深卷积层和小卷积核对视角等协变量因素具有较好的鲁棒性,图像细节信息提取的更丰富,最终得到较高的准确率。

1 步态能量图提取

1.1 步态轮廓图像提取

本文采用背景减除法^[17]提取运动中物体的轮廓。首先,对背景进行重构,然后,利用背景减除法检测出运动物体并提取其轮廓图像,采用二值化和形态学方法对提取出的轮廓图像进行处理,最后,对轮廓图像序列进行归一化操作。原始图像和归一化后步态轮廓图像如图 1 所示。



图 1 原始图像、二值图像和归一化步态轮廓图像

与人体生理特征(人脸、虹膜、指纹、掌纹)不同,人体步态具有周期性。步态周期一般定义为同一只脚两次落地的时间间隔。目前,步态周期检测中使用了一些常用的方法。Sarkar 等^[18]通过观察人体行走过程中身体下肢图像像素随时间的变化来确定步态周期。Culter 等^[19]使用自相关方法,根据轮廓随时间变化来提取步态周期,但该方法计算量大,不容易实现。因此本文通过观察人体行走过程中身体下肢图像像素随时间的变化计算得到步态周期。

1.2 步态能量图

步态能量图(GEI)是对步态周期进行时间归一化的累积能量图像,在图像中像素值越高表示这个位置运动更频繁,其能量更大,能反映其在一个周期内的行走状态。步态能量图像不仅保留了

原始图像的轮廓周期序列,相位变换,频率信息,还在很大程度上减少了步态周期的数量,降低了计算的复杂度。因此本文选择步态能量图对提取出的步态轮廓图像进行处理。对于步态周期 $f_i(x, y)$ 的给定二进制图像序列,步态能量图像计算表示为

$$G(x, y) = \frac{1}{N} \sum_{i=1}^N f_i(x, y) \quad (1)$$

式中: N 是完整步态周期的长度; i 表示图像序列; (x, y) 表示二维图像的像素坐标; 所得的灰度图 $G(x, y)$ 就是步态能量图,如图 2 所示。



图 2 归一化步态图像序列和合成步态能量图

2 基于 VGG 模型的步态识别

2.1 VGG 网络模型

2014 年,LeCun 提出了一种前馈神经网络被称为卷积神经网络,该网络无需对图像进行复杂的前期预处理操作,只需输入原始图像,所以在模式分类领域得到广泛的应用。卷积神经网络的基本结构包括两层,其一为卷积层也称特征提取层,每个神经元的输入连接前一层的局部感受野,并提取该局部的特征;其二是池化层也称特征映射层,网络的每个计算层由多个池化层组成,每个特征映射层是一个平面,平面上所有神经元的权值相等。因为一个映射面上的神经元共享权值,从而网络自由参数的个数减少了。卷积神经网络凭借其局部权值共享的特殊结构在图像处理和语音识别领域发挥着重要的作用,其结构与实际的生物神经网络非常相似。共享权值降低了网络的复杂性,且多维向量的图像可以直接输入网络这一特点降低了特征提取和分类识别中数据重建的复杂性。本文采用的是卷积神经网络中经典网络 VGG 模型,该模型基于 AlexNet 网络架构,其最大的特

点是通过 3×3 卷积的组合与堆叠, 提取输入图片更深层更抽象的特征^[20]。通过构建和组合多个 3×3 滤波器, 可以取得与 5×5 或 7×7 滤波器相同的计算结果, 连续小模板的滤波器比单个大的滤波器具有更好的非线性。同时 VGG 网络模型基于 AlexNet 网络模型对卷积层进行了加深, 增加了模型的运算复杂度, 模型的运算量由 720 M 上升约为 15 300 M, 但是 VGG 网络模型能获取到图像的高层语义信息, 更有利于提高步态识别的准确率。

2.2 网络模型训练

采用上述方法得到实验所需的步态能量图, 其大小为 240×240 的单通道图像, 即输入网络的数据维度为 $240 \times 240 \times 1$ 。将每张步态能量图作为一个 x 样本输入, 对应的行人 ID 作为标签 y 输入, 每个行人有相同张数的步态能量图, 将得到的步态能量图送入网络进行训练。VGG 网络模型是通过 100 万幅的 ImageNet 数据库中训练而成, 具

备较强的深度特征学习能力, 其拥有大量已训练完成的参数和权重, 特别是卷积层关于图像的曲线、边缘、轮廓的特征提取具备较强的能力^[21], 因此本文实验采用已经训练好的参数和权重, 再针对实际情况稍作改变, 主要是为了本文的实验效果。对于前 4 个卷积层, 每经过 2 个卷积层进行一次最大池化 (Max-pooling) 操作, 对于后面 9 个卷积层, 每经过 3 个卷积层进行一次最大池化操作, 所有卷积层链接的隐藏层都是用 tanh 作为激活函数。Dense 为通常所述的全连接层, 激活函数为 tanh。使用 softmax 层来分类, 分成 124 个类别。本文采用 dropout (0.5) 来防止过拟合现象, dropout 是按照一定概率将神经网络单元按照一定概率从网络中丢弃。为了使模型不断的优化, 选用的代价函数为对数损失函数 (binary_crossentropy)。优化器选择 SGD, 学习率为 0.01, 以 $1e-6$ 的速度衰减, 动量为 0.9。VGG 网络模型如图 3 所示。

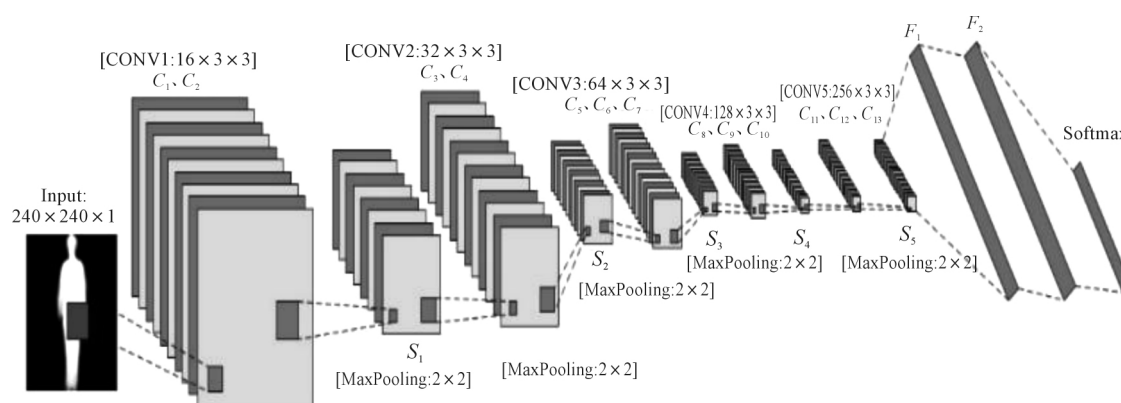


图 3 网络结构

3 实验结果与分析

3.1 实验平台及步态数据集

本文实验平台为 Google 公司公开的 TensorFlow^[22] 深度学习框架, 以及以其为后端的 Keras^[23] 高层神经网络 API, 两者均以 Python 为开发语言, 为加快深度学习模型的训练速度, 使用硬件平台为配置了英伟达 GPU 显存大小为 4G 的 Windows 系统。

实验采用的数据集是来自中科院的 CASIA 步态数据库 B, 在 CASIA B 中, 共有 124 个行人, 每个行人有两种变量, 即视角和步行场景。视角从 0° 到 180° 以 18° 为间隔均匀变化, 包含 11 个角度, 而步行场景有 3 个类别, 包括穿大衣 (cl-2)、背包 (bg-2)、正常行走 (nm-6)。在每种步行场景中又有 11 种角度变化, 因此共有 $124 \times 10 \times 11 \times 80 = 1\,091\,200$ 张图片。采用上述步态能量图提取方法, 对每个状态下的每个角度选择一个周期的

步态图像进行处理。尽管每个状态下的样本集略有区别,但是在对图像行处理时处理难度相差不大,因为都是在相同环境下行走,背景较为简单,而且行人的轮廓较为清晰,很容易获取。因此本文将 $124 \times 11 \times 10 \times 3 = 40\ 920$ 张步态能量图送入到网络进行训练识别。

当每个行人在场景中行走 10 次 (6 nm + 2 cl + 2 bg) 时,分别进行同状态实验和跨状态实验来验证算法的有效性。在同状态实验设计中,训练数据的行走场景和测试数据的行走场景相同,设置的实验(训练集 - 测试集)包括 (nm - nm)、(bg - bg)、(cl - cl)。在跨状态实验 (nm - bg) 和 (nm - cl) 设计中,将 nm (01 - 04) 4 个视频序列作为训练集,而剩下的 nm 和 bg、cl 作为测试集。详细实验设计如表 1。

表 1 CASIC B 数据集上的实验设计

参数	训练集	测试集
同状态	nm (01 - 04)	nm (05 - 06)
	bg - 01	bg - 02
	cl - 01	cl - 02
跨状态	nm (01 - 04)	nm (05 - 06)、bg (01 - 02)
	nm (01 - 04)	nm (05 - 06)、cl (01 - 02)

3.2 实验及结果

为了验证本文所采用方法的有效性,将使用准确率和平均 rank1 准确率来评估,平均 rank1 准确率为跨视角场景下在测试集上识别准确率的平均。迭代次数是指整个数据集通过网络的次数,每通过一次网络,网络便会学习数据集中的特征进行计算。本文实验迭代次数为 30 次,图 4 是 1 ~ 30 次迭代中在训练集和验证集上的准确率的变化图。

针对 CASIA B 数据库的实验结果如表 2,训练集和测试集均是在跨视角条件下进行实验,表 2 仅对行人行走场景进行描述,因为本文只考虑跨视角下的平均准确率,所以并未详细说明从 0° 到 180° 变化哪个视角步态识别效果最佳,对于所有视角使用平均 rank1 准确率。

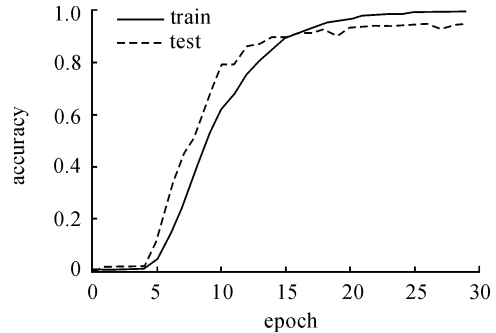


图 4 模型准确率

表 2 平均 rank1 准确率

参数	训练集	测试集	平均 rank1 准确率/%
同状态	nm	nm	95.5
	bg	bg	55.6
	cl	cl	55.3
跨状态	nm	bg	73%
	nm	cl	65.4%

由图 4 可看出,准确率在第 5 次迭代后剧烈上升,等到 15 次迭代后,准确率趋于平滑,上升较为缓慢,训练集准确率在前 15 次迭代均低于验证集准确率,在后 15 次迭代均大于。由表 2 明显可看出同状态的 nm - nm 的平均 rank1 准确率最高,能达到 95.5%,而在同状态 bg - bg、cl - cl 和跨状态 nm - bg、nm - cl 识别准确率都不是很高。在同状态 nm - nm 中,行人正常行走,只有视角变化,由此可得出步态能量能很好的表征步态特征,通过网络训练识别后准确率较高。而在同状态 bg - bg、cl - cl 和跨状态 nm - bg、nm - cl 中,行人分别穿着大衣和背着背包,同时伴随视角变化,由此可得出步态能量图未能充分的表征步态特征,导致识别准确率较低。主要是因为服装和携带物会对行人行走过程的轮廓存在遮挡,导致行走轮廓不能充分反映行人的行走特征,故从行走轮廓得到的步态能量图未能充分表征步态特征。因此本文所采用步态能量图和卷积神经网络相结合的步态识别方法对于正常行走过程中视角变化有较好的鲁棒性,相比于文献^[11]中的 GEINet 对于跨视角识别准确率为 91.6% 有了一点提升,主要是由于

VGG 网络比 GEINet 网络卷积层更深和利用了小卷积核的优点,通过 3×3 滤波器的组合与堆叠能提取输入图片更多更小的特征。但是对于服装、携带物等协变量因素鲁棒性较差,主要是因为服装和携带物会造成行人轮廓被遮挡,而步态能量图是由一个步态周期步态轮廓简单的平均导致步态能量图不能较好的表征被遮挡部分的特征,所以本文采用的步态能量图与 VGG 网络模型结合的步态识别方法对服装、携带物等协变量效果不佳。因此在下一步的研究中,将尝试使用其他的网络模型如 GoogLeNet、GAN 或者探索一种更能表征步态特征的步态模板来提取衣服及携带物无关的步态特征,提高步态识别率。同样的,也将尝试去改进本文提出的方法应用在更大型的步态数据库上,去探索一种识别率更高的步态识别方法,提高对衣服和携带物的鲁棒性。

4 结束语

近年来,深度学习被应用于图像处理各领域,因此本文尝试将深度学习用于步态识别,使用 VGG 卷积神经网络模型和步态能量图相结合的步态识别方法,在 CASIA 步态数据库 B 对该方法进行性能评估,发现该网络模型能够通过人类行走步态较为准确的识别出人类身份 ID,对视角变化有较好的鲁棒性。但是对于服装和携带物等情况,识别效果不佳,主要是因为行人轮廓存在遮挡导致步态能量图无法充分表征步态特征。因此在下一步的研究中,将尝试使用其他的网络模型如 GoogLeNet、GAN 或者探索一种更能表征步态特征的步态模板来提取衣服及携带物无关的步态特征,提高步态识别率。同样的,也将尝试去改进本文提出的方法应用在更大型的步态数据库上,去探索一种识别率更高的步态识别方法,提高对衣服和携带物的鲁棒性。

参考文献:

- [1] 何逸炜,张军平.步态识别的深度学习:综述[J].模式识别与人工智能,2018,31(5):442-452.
- [2] WU Zifeng, HUANG Yongzhen, WANG Liang, et al. A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 209-226.
- [3] LEE L, GRIMSON W E L. Gait analysis for recognition and classification[C]//Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, USA: IEEE, 2002.
- [4] BOBICK A F, JOHNSON A Y. Gait recognition using static activity-specific parameters[C]//Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001.
- [5] HU Rong. Human gait recognition[C]//HuaZhong University of Science and Technology Communication and Information Engineering PHD Thesis. Wuhan: 2010: 9-15.
- [6] BABAEE M, LI L, RIGOLL G. Person identification from partial gait cycle using fully convolutional neural networks[J]. Neurocomputing, 2019, 338(21): 116-125.
- [7] HE Yiwei, ZHANG Junping, SHAN Hongming, et al. Multi-task GANs for View-Specific Feature Learning in Gait Recognition[J]. IEEE Transactions on Information Forensics & Security, 2018: 102-113.
- [8] TOBY H W, LAM K H C, JAMES N K L. Gait flow image: A silhouette-based gait representation for human identification[J]. Pattern Recognition, 2011, 44(4): 973-987.
- [9] BASHIR K, XIANG T, GONG S. Gait recognition using Gait Entropy Image[C]//International Conference on Crime Detection & Prevention. IET, 2010.
- [10] IWAMA H, OKUMURA M, MAKIHARA Y, et al. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition[J]. IEEE Transactions on Information Forensics & Security, 2012, 7(5): 1511-1521.
- [11] Kohei Shiraga, Yasushi Makihara, Daigo Muramatsu, et al. GEINet: View-invariant gait recognition using a convolutional neural network[C]//International Conference on Biometrics. IEEE, 2016.
- [12] 周云成, 许童羽, 郑伟, 等. 基于深度卷积神经网络的番茄主要器官分类识别方法[J]. 农业工程学报,

- 2017, 33(15): 219–226.
- [13] 张建华, 孔繁涛, 吴建寨, 等. 基于改进 VGG 卷积神经网络的棉花病害识别模型[J]. 中国农业大学学报, 2018, 23(11): 161–171.
- [14] MANUEL L A, RUBEN G O, NICOLAI P. Appearance-invariant place recognition by discriminatively training a convolutional neural network [J]. Pattern Recognition Letters, 2017, 92: 89–95.
- [15] 杨亚楠, 夏斌, 袁文浩, 等. 应用卷积神经网络的测距算法研究[J]. 重庆理工大学学报(自然科学), 2018(3): 172–177.
- [16] 李耀龙, 张永科, 罗镇宝. 应用深度卷积神经网络的机场及机场内飞机目标识别技术[J]. 重庆理工大学学报(自然科学), 2018(3): 210–216.
- [17] 刘仲民, 何胜蛟, 胡文瑾, 等. 基于背景减除法的视频序列运动目标检测[J]. 计算机应用, 2017, 37(6): 1777–1781.
- [18] SARKAR S, PHILLIPS P J, LIU Z, et al. The humanID gait challenge problem: data sets, performance and analysis [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(2): 162–177.
- [19] CUTLER R. Robust real-time periodic motion detection: analysis and applications [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(8): 781–796.
- [20] LIU H, LU J. Group-aware deep feature learning for facial age estimation [J]. Pattern Recognition, 2017, 66(C): 82–94.
- [21] FENG L, PO L M, LI Y. Integration of image quality and motion cues for face anti-spoofing: Aneural network approach [J]. Journal of Visual Communication & Image Representation, 2016, 38(23): 451–460.
- [22] TensorFlow. 谷歌深度学习框架 [EB/OL]. [2018–04–15] <https://www.tensorflow.org/?hl=zh-cn>.
- [23] KERAS. 基于 Python 的深度学习库 [EB/OL]. [2018–06–05] <https://keras.io/zh/>.

(责任编辑 何联格)