



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象学报 中国图形学报

2022
03
VOL.27

ISSN1006-8961
CN11-3758/TB



医学图像及临床应用



第27卷第3期（总第311期）
2022年3月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

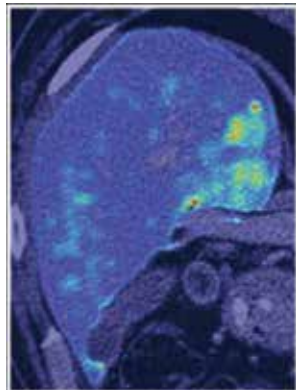
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

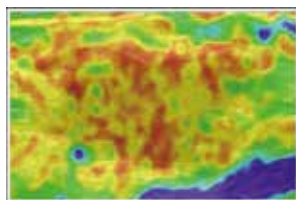
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



慢阻肺患者CT图像中肺内血管分割及定量分析(第0762页)



基于特征选择与残差融合的肝肿瘤分割模型(第0838页)



面向乳腺超声计算机辅助诊断的两阶段深度迁移学习(第0898页)

专刊封面展示了上海科技大学生物医学工程学院和上海联影智能医疗科技有限公司合作的脑影像智能分析方向研究成果。

《中国图象图形学报》医学图像及临床应用专刊简介

沈定刚, 刘天明, 周涛, 夏勇, 白相志, 王乾, 李刚, 徐寿平, 刘昊, 韩向娣 0653

综述

中国医学影像人工智能20年回顾和展望

蒋希, 袁奕宣, 王雅萍, 肖振祥, 朱美芦, 陈泽华, 刘天明, 沈定刚 0655

迁移学习在医学图像分类中的研究进展

黎英, 宋佩华 0672

生成对抗式网络及其医学影像应用研究综述

张颖麟, 胡衍, 东田理沙, 刘江 0687

基于深度学习的心脏磁共振影像超分辨率前沿进展

李书林, 冯朝路, 于鲲, 刘鑫, 江鑫, 赵大哲 0704

CT图像肺及肺病变区域分割方法综述

冯龙锋, 陈英, 周滔辉, 胡菲, 易珍 0722

计算机断层扫描图像

结构图注意力网络的新冠肺炎轻重症诊断

刘彦北, 李赫南, 张长青, 肖志涛, 张芳, 陈英, 高耀宗, 石峰, 单飞, 沈定刚 0750

慢阻肺患者CT图像中肺内血管分割及定量分析

赵宏, 李璋, 张杰华, 王琨, 孙家兴, 廖玺铭, 阎昱升, 钟正, 张鑫, 孙健, 于起峰, 葛俊辉 0762

基础信息保持和细节强化的胸部CT图像增强

张余, 张顺利, 白相志, 张利 0774

面向小样本股骨骨折分型的多视角注意力融合方法

张亚东, 汪玲, 兰海, 翟禹樵, 程洪 0784

三维多尺度嵌套U结构CT影像肺结节检测

许正玺, 张少敏, 支力佳, 周涛 0797

融合多尺度残差和注意力机制的特发性肺纤维化进展预测

陈舞, 孙军梅, 李秀梅 0812

COVID-19肺部CT图像多尺度编解码分割

陆倩杰, 柏正尧, 樊圣澜, 周雪, 许祝 0827

基于特征选择与残差融合的肝肿瘤分割模型

乔伟晨, 黄冕, 刘利军, 黄青松 0838

磁共振图像

面向多模态MRI脑胶质瘤区域三维分割与生存期预测的级联U-Net网络

余力, 刘霄雪, 闫朝阳, 李建瑞, 张志强, 黄轲, 徐军 0850

心脏动态MRI图像分割的时空多尺度网络

徐佳陈, 肖志勇 0862

融合跨阶段深度学习的脑肿瘤MRI图像分割

夏峰, 邵海见, 邓星 0873

融入平滑组稀疏化的脑部MRI图像分类

黄帅辉, 王金凤 0885

超声图像

面向乳腺超声计算机辅助诊断的两阶段深度迁移学习

贡荣麟, 施俊, 周玮珺, 汪程 0898

结合注意力机制的乳腺双模态超声分类网络

赵绪, 龚勋, 樊琳, 罗俊 0911

中医图像

视觉Transformer与多特征融合的脑卒中检测算法

赵琛琦, 王华虎, 赵涓涓, 冀伦文, 王麒达, 李慧芝, 赵紫娟 0923

多分支深度特征融合的中医脑卒中辅助诊断

王麒达, 冀伦文, 强彦, 王华虎, 赵琛琦, 李慧芝, 赵紫娟 0935

研究应用

多区域融合注意力网络模型下的核性白内障分类

章晓庆, 肖尊杰, Risa Higashita, 陈婉, 胡衍, 袁进, 刘江 0948

面向高度近视条纹损伤的深监督特征聚合网络

谭晓, 刁逸超, 陈新建, 石霏, 樊莹, 谢嘉受, 朱伟芳 0961

结合目标检测与匹配修正的手腕骨兴趣区域提取

毛科技, 汪敏豪, 陈立建, 陆伟, 武坤秀, 陈庆章, 赵小敏 0973

面向旷场实验视频分类的特征拼接矩阵学习方法

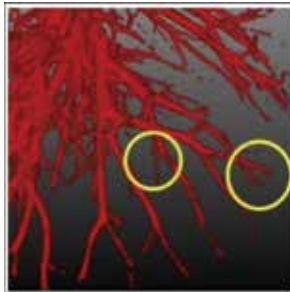
肖照林, 杨志林, 刘欢, 金海燕 0988

基于图像的介入器械前端变形建模与受力估计

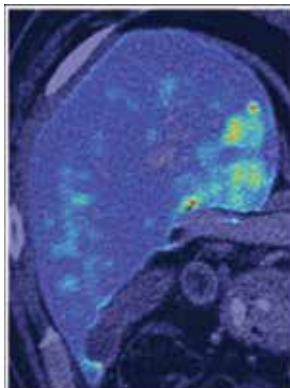
任龙飞, 孟偲, 刘博, 白相志, 周付根 1001

CONTENTS

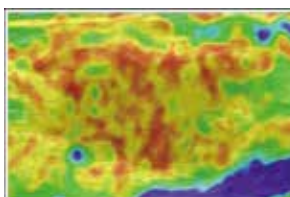
JOURNAL OF IMAGE AND GRAPHICS



Segmentation and quantitative analysis of intrapulmonary vasculature in CT images from COPD patients(P0762)



Feature selection and residual fusion segmentation network for liver tumor(P0838)



Two-stage deep transfer learning for human breast ultrasound computer-aided diagnosis(P0898)

Review

- A 20-year retrospect and prospect of medical imaging artificial intelligence in China
Jiang Xi, Yuan Yixuan, Wang Yaping, Xiao Zhenxiang, Zhu Meilu, Chen Zehua, Liu Tianming, Shen Dinggang 0655
- Review of transfer learning in medical image classification
Li Ying, Song Peihua 0672
- A review of generative adversarial networks and the application in medical image
Zhang Yinglin, Hu Yan, Risa Higashita, Liu Jiang 0687
- Critical review of human cardiac magnetic resonance image super resolution reconstruction based on deep learning method
Li Shulin, Feng Chaolu, Yu Kun, Liu Xin, Jiang Xin, Zhao Dazhe 0704
- Review of human lung and lung lesion regions segmentation methods based on CT images
Feng Longfeng, Chen Ying, Zhou Taohui, Hu Fei, Yi Zhen 0722

Computerd Tomography Image

- Diagnosis of COVID-19 by using structural attention graph neural network
Liu Yanbei, Li Henan, Zhang Changqing, Xiao Zhitao, Zhang Fang, Wei Ying, Gao Yaozong, Shi Feng, Shan Fei, Shen Dinggang 0750
- Segmentation and quantitative analysis of intrapulmonary vasculature in CT images from COPD patients
Zhao Hong, Li Zhang, Zhang Jiehua, Wang Kun, Sun Jiaxing, Liao Ximing, Yan Yusheng, Zhong Zheng, Zhang Xin, Sun Jian, Yu Qifeng, Ge Junhui 0762
- Human chest CT image enhancement based on basic information preservation and detail enhancement
Zhang Yu, Zhang Shunli, Bai Xiangzhi, Zhang Li 0774
- Multi-view attention fusion method for few-shot femoral fracture classification
Zhang Yadong, Wang Ling, Lan Hai, Zhai Yuqiao, Cheng Hong 0784
- Detection of pulmonary nodules in three-dimensional multiscale nested U-structure computed tomography images
Xu Zhengxi, Zhang Shaomin, Zhi Lijia, Zhou Tao 0797
- Multi-scale residual and attention mechanism fusion based prediction for the progression of idiopathic pulmonary fibrosis
Chen Wu, Sun Junmei, Li Xiumei 0812
- Multiscale codec network based CT image segmentation for human lung disease derived of COVID-19
Lu Qianjie, Bai Zhengyao, Fan Shenglan, Zhou Xue, Xu Zhu 0827
- Feature selection and residual fusion segmentation network for liver tumor
Qiao Weichen, Huang Mian, Liu Lijun, Huang Qingsong 0838

Magnetic Resonance Image

- Brain glioma tumor segmentation and survival prediction from multi-modality MRIs via cascaded U-Net
Yu Li, Liu Xiaoxue, Yan Chaoyang, Li Jianrui, Zhang Zhiqiang, Huang Yunzhi, Xu Jun 0850
- A spatio-temporal multi-scale network for cardiac dynamic MRI image segmentation
Xu Jiachen, Xiao Zhiyong 0862
- Cross-stage deep-learning-based MRI fused images of human brain tumor segmentation
Xia Feng, Shao Haijian, Deng Xing 0873
- Brain MRI image classification incorporating smooth group sparseness
Huang Shuaihui, Wang Jinfeng 0885

Ultrasound Image

- Two-stage deep transfer learning for human breast ultrasound computer-aided diagnosis
Gong Ronglin, Shi Jun, Zhou Weijun, Wang Cheng 0898
- Attention-based networks of human breast bimodal ultrasound imaging classification
Zhao Xu, Gong Xun, Fan Lin, Luo Jun 0911

Traditional Chinese Medical Image

- Cerebral stroke detection algorithm for visual Transformer and multi-feature fusion
Zhao Chenqi, Wang Huahu, Zhao Juanjuan, Ji Lunwen, Wang Qida, Li Huizhi, Zhao Zijuan 0923
- Multi-branch deep feature fusion method for traditional Chinese medicine intervened Human cerebral stroke aided diagnosis
Wang Qida, Ji Lunwen, Qiang Yan, Wang Huahu, Zhao Chenqi, Li Huizhi, Zhao Zijuan 0935

Research and Application

- Nuclear cataract classification based on multi-region fusion attention network model
Zhang Xiaoqing, Xiao Zunjie, Risa Higashita, Chen Wan, Hu Yan, Yuan Jin, Liu Jiang 0948
- Deep-supervision and feature-aggregation network for linear lesion segmentation of high myopia
Tan Xiao, Diao Yichao, Chen Xinjian, Shi Fei, Fan Ying, Xie Jiamin, Zhu Weifang 0959
- Hand-wrist region of interest extraction based on object detection and matching correction
Mao Keji, Wang Minhao, Chen Lijian, Lu Wei, Wu Kunxiu, Chen Qingzhang, Zhao Xiaomin 0973
- A spliced feature matrices learning method for open field test video classification
Xiao Zhaolin, Yang Zhilin, Liu Huan, Jin Haiyan 0988
- Image-based deformation modeling and force estimation of the front end of vascular interventional surgical apparatus
Ren Longfei, Meng Cai, Liu Bo, Bai Xiangzhi, Zhou Fugen 1001

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2022)03-0988-13

论文引用格式: Xiao Z L, Yang Z L, Liu H and Jin H Y. 2022. A spliced feature matrices learning method for open field test video classification. Journal of Image and Graphics, 27(03): 0988-1000 (肖照林, 杨志林, 刘欢, 金海燕. 2022. 面向旷场实验视频分类的特征拼接矩阵学习方法. 中国图象图形学报, 27(03): 0988-1000) [DOI: 10.11834/jig.210606]

面向旷场实验视频分类的特征拼接矩阵学习方法

肖照林^{1,2}, 杨志林¹, 刘欢¹, 金海燕^{1,2*}

1. 西安理工大学计算机科学与工程学院, 西安 710048; 2. 陕西省网络计算与安全技术重点实验室, 西安 710048

摘要: **目的** 旷场实验(open field test, OFT)是行为学与药理实验分析中常用的实验方法。为了对比测试组和参考组被测小鼠的行为特征差异,通常需要耗费大量精力对旷场实验数据进行处理和观测。由于旷场实验数据量大且较依赖观测人员的主观判断,导致对小鼠行为差异观测的精度较低且缺乏量化评价指标。为此,本文提出一种基于卷积神经网络(convolutional neural networks, CNN)的旷场实验视频分类方法,可基于量化特征对两组小鼠的行为差异自动分类。**方法** 从视频空域和时域中提取22维的小鼠运动行为特征,经过量化后生成特征矩阵,进而以矩阵拼接方式构造可学习的行为特征矩阵样本,利用不同结构卷积神经网络对提取的行为特征矩阵样本进行训练和分类,并分析网络结构对分类结果的影响,在实现两组小鼠分类的基础上,对不同维度小鼠行为特征对分类精度的重要性进行评价。**结果** 在真实旷场实验数据集上的实验分析表明,本文算法的分类准确率为99.25%。此外,由实验结果分析发现小鼠的大角度转向频次、停留区域与时间对小鼠分类的重要性高于其他维度特征。**结论** 提出的特征拼接矩阵学习方法能够准确识别两组小鼠旷场实验视频的差异,本文方法的分类准确率明显优于现有人工分析及经典机器学习方法。

关键词: 旷场实验(OFT);行为分析;视频分类;特征拼接矩阵;卷积神经网络(CNN)

A spliced feature matrices learning method for open field test video classification

Xiao Zhaolin^{1,2}, Yang Zhilin¹, Liu Huan¹, Jin Haiyan^{1,2*}

1. Department of Computer Science, Xi'an University of Technology, Xi'an 710048, China;

2. Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an 710048, China

Abstract: **Objective** open field test (OFT) video classification is a wide spread method to assess locomotors activity and mice habits control in pharmacological experiments. One major goal of traditional OFT is to sort out distinguishable features between a testing group and a reference group. Researchers will analyze those video clips to find out animal behaviors differences for the two groups of mice based on digital OFT videos records. The manual inspection of OFT data is time consumed and high cost involved. A detailed analysis process relies heavily on professional experiences. Our research illustrates a learning-based classification through an OFT-videos-based quantitative feature extraction. We show that a spliced feature matrix learning method has its priority than current independent features. Hence, a convolutional neural network based solution sort out the oriented video clips of two groups of mice. Our analyzed results can be summarized as following:

收稿日期: 2021-07-26; 修回日期: 2021-11-18; 预印本日期: 2021-11-25

* 通信作者: 金海燕 jinhaiyan@xaut.edu.cn

基金项目: 国家自然科学基金项目(61871319)

Supported by: National Natural Science Foundation of China(61871319)

1) a novel spliced feature matrix of OFT video is demonstrated based on 22-dimension quantitative behavioral features.

2) To analyze the influence of network structure on the classification, we design and test 8 different convolutional neural networks in the experiments. **Method** Our research focuses on the varied characteristics between the two groups of mice macroscopic behavior. In the context of the spatial and temporal domain of OFT video data, 22 distinct types of features are extracted including average crawling speed and distance, the positional distribution of staying, resting time, turning frequencies and so on. These quantitative vector-based descriptions of the OFT video can be well distinguished based on traditional machine learning classifiers, such as K-means clustering, boosting classification, support vector machine (SVM), or convolutional neural network (CNN). Some quantitative feature vectors are proposed to separate the testing group from the reference group. Another critical factor is the mice crawling path, which cannot be vector-oriented. A novel regularization and fusion are employed to illustrate both the quantitative features and non-quantitative crawling path. By constructing the targeted self-correlation matrix of weighted feature vectors, a 484×484 quantitative feature matrix is spliced with a 484×484 crawling path image, and it leads a spliced feature matrix representation with a 2D dimension of 484×968 . We introduce a CNN-based classification of spliced feature matrices from different network structures. Based on 4-layers CNN to 10-layers CNN, the quantitative feature matrix or the spliced feature matrices are learned via the basic network structures, which are we evaluate the impacts of different network structures and feature dimensions on the precision of OFT video classification further. In the experiments, we test the demonstrated feature extraction and classification on the real OFT dataset. The dataset is composed of OFT video of 32 mice, including 17 mice in the test group and 15 mice in the reference group. The testing animal group is injected with a kind of anti-depressants, and the reference group injected with only placebo. Each mouse is eye-captured independently using a video camera for a 24 hours timescale detection in the OFT experiment. These video data is cropped into targeted short duration video clips, each of which is 10 minutes long. We construct a dataset based on 3 080 testing samples and 1 034 reference samples. In the training analysis, the training dataset is based on 3 000 samples and the testing dataset is based on 1 114 samples. **Result** Our demonstration indicates that the proposed algorithm is optimized to manual classification and supported-vector-machine solution. In respect of the experiment datasets, the classification precision of the proposed algorithm is up to 99.25%. The accurate classification of mouse OFT video can be achieved via a simplified network structure, such as a 9-layers CNN. The varied feature dimension to classification accuracy is not balanced. In terms of the ablation results among different quantitative features, the large angle turning, staying time and positional distribution are critical for identifying its grouping than other quantitative feature dimensions. Non-quantitative crawling path image has obvious contribution to classification, which improves 2%—3% precision in classification. The demonstration also verifies can achieve good classification via a sample network structure. **Conclusion** Overall, we propose a CNN-based solution to classify the testing group OFT videos from the reference group of OFT videos via the spliced feature matrix learning. Both the quantitative features and qualitative crawling path are melted into a novel regularization and fusion. The identified classification has optimized on manual classification and traditional SVM methods. Our classification precision can reach 99.25% on the experimental dataset. It demonstrates a great potential for the OFT analysis. The further research work can verify the relationship between distinguishable features and detailed mouse behavior. The learning network optimization can be further called to enhance the generalization of the OFT classification.

Key words: open field test (OFT); behavioral analysis; video classification; spliced feature matrix; convolutional neural network (CNN)

0 引言

旷场实验(open field test, OFT) (Hall, 1934) 又称敞箱实验, 是评价实验动物在新异环境中自主行为、探究行为与紧张度的一种方法, 在医学、药理学以及行为学研究中广泛应用。旷场实验将实验组和

参考组的小鼠分别放在具有相同环境的独立实验箱中观察, 记录并分析两组小鼠的运动时间、运动距离、停留区域及停留时间等信息, 从而评价小鼠受到药物及外界刺激后兴奋、焦虑和恐惧等状态的变化。目前, 旷场实验仍是用于评价动物自发活动及焦虑症状的经典行为学实验 (Walsh 和 Cummins, 1976)。医学研究人员在分析旷场实验时对小鼠行为特征常

依据观察和人工经验进行提取和统计,导致可分析特征的维度较低,且对两组小鼠特征差异分析及判断较为依赖主观经验。因此,旷场实验人工分析及数据统计面临精确性与可靠性的双重挑战。

本世纪以来,图像分割、目标检测和跟踪、数据统计与分析等方法有效推动了视频分析领域的快速发展。OFT 视频分析技术在人工智能与模式识别领域受到关注(Stanford,2007;Nováková 等,1997),深度学习技术在目标分类、特征识别以及图像和视频处理领域取得了众多突破性进展。结合深度学习理论,本文探索小鼠 OFT 视频运动特征的量化方法以及分类问题,采用拼接矩阵方式形成一种对 OFT 视频中小鼠运动特征的矩阵化表示,进而结合卷积神经网络(convolutional neural network, CNN)的 Tensorflow 框架,提出一种能够有效区分测试组和参考组小鼠视频的深度学习视频分类方法。该方法有助于量化分析 OFT 视频中小鼠的行为差异,为 OFT 分析提供自动化分类模型及计算方法。本文创新可总结为:1)提出将 OFT 视频中小鼠的可量化特征进行矩阵化表示,并与难以量化的运动路径图像进行拼接,形成对 OFT 视频运动特征的特征拼接矩阵,形成可用于特征学习的 OFT 视频量化数据。2)基于 OFT 视频的特征拼接矩阵,提出基于卷积神经网络的小鼠分类方法,并在真实 OFT 数据集上验证了分类的准确性。在此基础上,进一步分析不同卷积神经网络结构的影响,发现仅采用层数较少的简单网络便可实现对特征拼接矩阵高精度分类。3)本文实验均基于真实 OFT 数据,在受过训练的被试者很难分辨两组小鼠视频差异的情况下,本文算法仍可准确区分两组小鼠视频之间的差异。这为基于 OFT 视频的小鼠行为差异量化分析提供了新方法,并为 OFT 视频分类提供了高精度的自动化方案。

1 相关工作

1.1 基于机器学习的视频分类技术

自 20 世纪 90 年代 Cortes 和 Vapnik(1995)采用软边距非线性支持向量机(support vector machine, SVM)实现高精度的手写字符识别以来,基于 SVM 的目标分类方法突破了传统特征建模类方法的局限,在多种图像分类和目标识别任务中取得了优异的性能(Durand 等,2015;Schuldt 等,2004;Lin 和

Hauptmann,2002),谷军霞等人(2009)对人类行为分析相关早期工作进行了总结。支持向量机(SVM)是一种监督式学习的方法,主要针对小样本的数据分类(陆波 等,2005),且在多分类场景中效果不理想。

近 10 年,以卷积神经网络为代表的深度学习技术在众多领域得到了广泛应用(Deng 和 Yu,2014;Nair 等,2015;Nam 和 Han,2016)。在视频分类领域中,Karpathy 等人(2014)提出了基于 CNN 的多分辨率视频分类方法。Ng 等人(2015)将循环神经网络和长短期记忆网络连接到底层的 CNN 输出端,使视频分类的精度上又有了进一步提升。Xu 等人(2016)提出了递归卷积神经网络(recurrent convolutional neural network, RCNN),利用帧图像中的局部特征、密集特征以及连续帧之间的时间特征对视频进行分类,对不同种类的视频分类具有显著效果。Wang 等人(2016)提出了一个 CNN-RNN 组合框架,以解决多标签图像分类问题。Zhou 等人(2000)将视频的特征提取和内容分析相结合,提出了一种无监督的视频分类系统,并在视频语义分类中取得了较好效果。冯文刚等人(2011)采用无监督学习方式显示了对无线胶囊内窥镜视频的准确分类。Carreira 和 Zisserman(2017)提出了一种采用 3D 卷积神经网络构建的双流融合视频分类方法,并在其发布的 Kinetics 数据集上取得了准确的人类行为视频分类结果。Jiang 等人(2018)在研究了诸如人类的复杂行为之后,提出了一种新的视频分类框架,联合利用特征关系和类关系提高分类性能,提出的正则化语义视频建模对视频分类在当前的视频分类中具有明显优势。Brattoli 等人(2020)设计端到端的 3D 卷积神经网络学习待分类样本内部的视觉特征,可由静态图像预训练分类模型,并在指定视频数据上进行微调(fine-tune),已获得更好的分类性能。

1.2 基于机器学习的旷场实验数据处理技术

Choleris 等人(2001)采用旷场实验研究药物使用与小鼠焦虑行为的关系,使用实验小鼠各种行为(即速度、路程、不同区域的停留时间以及站立次数等)的持续时间量化测试药物的药理性能。基于视频的自动化程序处理技术,Decker 和 Hamprecht(2014)提出精确对小鼠身体不同部分进行检测的方法,有助于对小鼠行为进行识别和分类。Monteiro 等人(2014)提出了一种基于深度映射的

小鼠图像分割方法,并给出了 4 种用于小鼠动作识别和分类的典型特征。此外,Hong 等人(2015)尝试对两只小鼠进行更复杂的社会交互行为的识别,采用基于视频的模式识别方法可有效区分 13 种不同的小鼠行为。

在机器学习理论发展的同时,Burgos-Artizzu 等人(2012)使用自适应提升(adaptive boosting, Ada-Boost)方法(Freund 和 Schapire, 1997)作为集成分类器对实验小鼠进行了社会行为分析,对比了轨迹特征、时空特征以及两者的特征拼接矩阵在 Ada-Boost 下的识别率,结果显示特征拼接矩阵的识别率明显优于单独使用其他两种特征。Giancardo 等人(2013)构建了一种实验小鼠的行为识别系统,用时间随机森林法(Svetnik 等,2003)对小鼠的交互行为进行分类。Katsageorgiou 等人(2016)提出一个数据驱动模型的无监督学习方法对小鼠的相互影响进行分析。上述小鼠行为学分析仍采用单一特征或少量特征组合方式对一只或多只小鼠进行检测和分类,检测精度以及特征可分性仍存在可优化空间。

本文介绍一种面向旷场实验视频分类的特征拼接矩阵学习算法,采用 CNN 作为基本分类器,通过对 22 维小鼠量化特征所拼接的特征矩阵学习,实现

高精度的旷场实验小鼠视频准确分类。

2 基于视频分析的小鼠视频分类

本文 OFT 实验中的小鼠分为参考组和测试组,在视频拍摄前对测试组小鼠注射某种抗抑郁药物,对参考组小鼠仅注射安慰剂(不含药物的试剂)。本文目标是采用深度学习方法对测试组和参考组小鼠视频进行自动分类,同时量化评价 OFT 视频中小鼠各个运动特征的可分类特性。

针对小鼠的 OFT 视频数据,本文提出一种基于特征拼接矩阵学习的 CNN 分类方法,算法流程如图 1 所示。图 1 中的视频分类流程大致为,从视频集中提取多种小鼠运动特征,并将不同特征进行量化计算,再对量化后的特征进行组合及向量化。在此基础上,将向量化的特征进行矩阵化表达和特征拼接生成特征矩阵 $\mathbf{B}$ 。在此基础上,本文提出一种可学习特征矩阵的卷积神经网络,实现对小鼠视频的准确分类。本文方法不需要将视频中的每一帧图像都作为卷积神经网络的输入,从而显著减少神经网络训练的参数量和计算量,因而具有较好的计算性能。

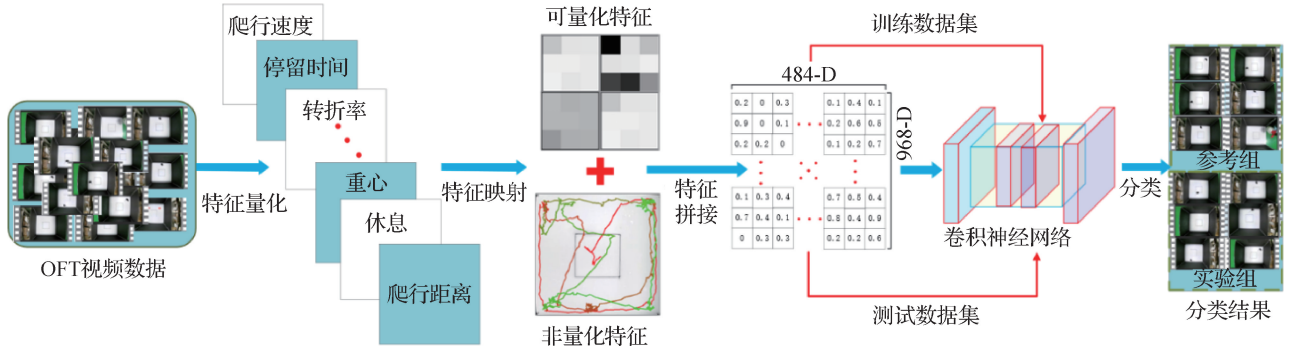


图 1 基于特征拼接矩阵的小鼠视频分类流程

Fig. 1 An overview of the proposed OFT video classification based on the spliced feature matrices

2.1 OFT 视频小鼠行为的特征提取方法

为了量化小鼠的行为特征,本文提取小鼠的爬行总体距离 s 、爬行速度 v 、停留时间 A_i 、转折率 $f_{(i,k)}$ 与休息时间 t_r 。假设相邻视频帧间的时间为 Δt ,小鼠在一段时间内的距离可由积分 $s = \int_0^t v dt$ 进行计算,其中 v 为小鼠在 Δt 内的瞬时速度, $\bar{v}$ 表示小鼠运动的平均速度。由此,小鼠 OFT 视频的可量化特征信息集合为 $\{\bar{v}, s, A_i, t_r, f_{(i,k)}\}$,这些可量化的

特征可由一个 22 维的数值化特征向量进行表示。

本文采用特征矩阵 $\mathbf{A}$ 描述小鼠在旷场不同区域的停留时间,具体为

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & A_3 \\ A_4 & A_5 & A_6 \\ A_7 & A_8 & A_9 \end{bmatrix} \quad (1)$$

式中, $\mathbf{A}$ 表示小鼠在旷场中停留总时间的区域划分, $\{A_i | i = 1, 2, \dots, 9\}$ 为小鼠在旷场不同区域的

停留时间。

为了求取 A_i , 本文利用小鼠重心所属 A_i 的范围来统计小鼠在此区域内的停留时间。计算过程为将小鼠在某时刻在 A_i 区域内的停留时间记为 t' , 当小鼠重心再次属于 A_i 时更新 $A_i = t' + \Delta t$ 。

为了计算小鼠爬行的转折率 $f(I_i)$, 对一段 OFT 视频帧集合 $\{I_1, I_2, \dots, I_n\}$, 本文取前后相邻间隔为 k 的两帧图像, 记为 I_{i-k} 和 I_{i+k} 。对 3 帧图像 I_{i-k} 、 I_i 和 I_{i+k} 计算其中小鼠的重心, 计算过程如图 2 所示。

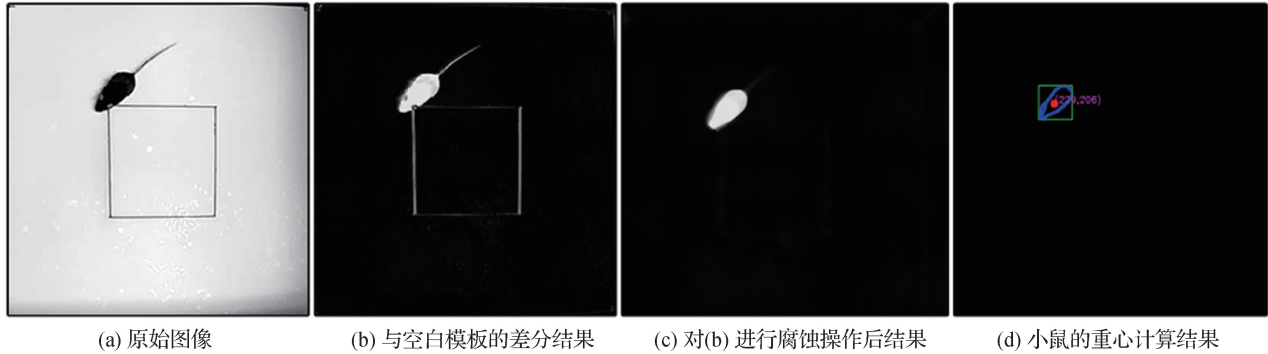


图 2 求小鼠重心过程

Fig. 2 The calculation of the gravity center ((a) original image; (b) difference between the two templates; (c) results of corrosion operation; (d) calculation result of the mouse gravity center)

采用图像的 1 阶矩计算小鼠重心, 计算为

$$x_c = \frac{M_{10}}{M_{00}}, y_c = \frac{M_{01}}{M_{00}} \quad (2)$$

$$\begin{cases} M_{00} = \sum_i \sum_j V(i, j) \\ M_{10} = \sum_i \sum_j i \cdot V(i, j) \\ M_{01} = \sum_i \sum_j j \cdot V(i, j) \end{cases} \quad (3)$$

式中, M_{00} 表示图像的 0 阶矩, 可以用来求取图像的轮廓和连通区域的面积。 M_{10} 和 M_{01} 分别表示图像在 x 和 y 方向的 1 阶矩, $V(i, j)$ 为二值图像中 (i, j) 位置像素点值。小鼠重心的计算结果如图 2(d) 所示。通过式(2)(3)可得 I_{i-k} 、 I_i 和 I_{i+k} 3 帧图像中小鼠的重心分别为 $p_{i-k}(x_c^{i-k}, y_c^{i-k})$ 、 $p_i(x_c^i, y_c^i)$ 和 $p_{i+k}(x_c^{i+k}, y_c^{i+k})$ 3 点, 其中 (x_c^{i-k}, y_c^{i-k}) 、 (x_c^i, y_c^i) 和 (x_c^{i+k}, y_c^{i+k}) 为相邻 k 帧的小鼠重心坐标。则在 i 时刻的小鼠爬行转角 $f_{(i,k)}$ 的计算为

$$f_{(i,k)} = \arccos\left(\frac{p_{i-k} \cdot p_i \cdot p_{i+k}}{\|p_{i-k} - p_i\| \|p_i - p_{i+k}\|}\right) \quad (4)$$

对于小鼠的休息时间 t_r , 本文统计相邻 k 帧中图像中重心平均移动距离 $l < l_{\min}$ 的总时间, $l_{\min}$ 为最小不计运动距离的长度阈值。

$$l = \frac{1}{k} \sum_{j=0}^k \|p_{i+j} - p_i\| \quad (5)$$

2.2 行为特征矩阵的拼接生成

依据 2.1 节所述, 本文提取小 OFT 视频特征集合 $\{\bar{v}, s, A_i, t_r, f_{(i,k)}\}$, 包括平均速度 $\bar{v}$ 、爬行总体距离 s 、休息时间 t_r 、 A_1 — A_9 的区域停留时间和 f_1 — f_{11} 的转角频次信息等 22 维的可量化数值化。通过分析大量小鼠 OFT 视频的量化特征数据, 发现不同的特征维度间存在相关性。例如, 分区域停留时间 A_1 — A_9 之间存在明显空间关联性, 爬行总体距离与休息时间存在反相关特点, 爬行速度与角度频次也存在相关性。因此, 采用 484×484 维特征相关矩阵作为每段视频的特征表达。为了进一步提高 OFT 视频分类质量, 本文将可量化特征矩阵与小鼠路径图像进行拼接, 作为分类学习网络的输入以提高分类性能。

与现有基于深度学习的视频分类方法不同, 本文采用可量化特征与非量化特征结合的方式解决 OFT 视频的分类问题。在特征量化阶段, 为了更好地识别小鼠直线前进和掉头行为, 将 $0^\circ \sim 180^\circ$ 可能的转弯范围划分为 11 个转角区间统计小鼠爬行中转弯的频率和幅度。除了最大和最小区间角度范围为 $0^\circ \sim 25^\circ$ 和 $160^\circ \sim 180^\circ$ 以外, 其余区间均为 15° 范围。由此, 采用 22 维特征向量 F 量化表示小鼠 OFT 视频特征, 即

$$F = [\bar{v}, A_1, \dots, A_9, t_r, f_1, \dots, f_{11}] \quad (6)$$

在此基础上,所有小鼠视频可采用经典 SVM 进行特征分类。为利用深度学习处理方法同时处理小鼠可量化特征以及非量化的小鼠爬行路径图像 I_{Trk} ,对可量化特征 F 及小鼠路径图像采用矩阵拼接,构建小鼠行为特征矩阵 B ,具体为

$$B = G(F_w F_w^T) G(F_w F_w^T)^T \circ H(I_{\text{Trk}}) \quad (7)$$

式中, $F_w = \omega F$ 代表对特征维度的加权后向量,其中权重系数为 ω ,由不同维度对 OFT 视频分类的重要性排序决定。 $G(\cdot)$ 代表将矩阵按行展开成向量, $H(I_{\text{Trk}})$ 为小鼠的运动路径信息, $H(\cdot)$ 为对路径图像 I_{Trk} 的尺度变换,“ $\circ$ ”代表两个矩阵的拼接。

计算 $F_w F_w^T$ 可得加权特征向量各维度的一阶自相关矩阵,大小为 22×22 维。在此基础上,构建特征间更为复杂的二阶相关性矩阵。至此,将 OFT 小鼠行为特征及其二阶相关性用一个维度为 484×484 的矩阵进行表示。随后,采用矩阵拼接方式,将非量化的小鼠爬行路径图像与可量化的运动特征信息进行拼接,形成 484×968 维度的小鼠运动特征

拼接矩阵。

对于测试组和参考组小鼠的大量视频,本文对每段视频进行特征矩阵样本生成,由此将两组小鼠的视频分类问题转变为对上述特征矩阵集合 $Q = \{B_1, B_2, \dots, B_m\}$ 的二分类问题。综上所述,对量化特征采用其二阶相关性的矩阵化表达,结合非量化特征构建对 OFT 视频的特征拼接矩阵,并采用卷积神经网络对此类数值矩阵进行训练,以获取对测试组和参考组 OFT 视频的准确分类。

2.3 分类方法

为了对小鼠行为特征矩阵进行分类,采用卷积神经网络 CNN 作为自动分类器,结构如表 1 所示。整个卷积过程中,所有卷积层均采用 3×3 的卷积核进行构建。为评估 CNN 网络层数以及参数量对分类结果的影响,分别采用了 8 种不同的网络结构(见表 1),其中 A—D 4 种网络结构仅对可量化特征矩阵进行计算,E—H 网络结构对可量化特征矩阵与小鼠路径图像的拼接矩阵进行计算。

表 1 本文采用的卷积神经网络结构

Table 1 The structures of convolutional neural networks in this paper

网络	层数	输入	结构							
A	4	484×484	卷积层	卷积层	—	—	—	—	—	卷积层 全连接层 softmax 层
B	6	484×484	卷积层	卷积层	下采样层 * 2	卷积层	—	—	—	卷积层 全连接层 softmax 层
C	8	484×484	卷积层	卷积层	下采样层 * 2	卷积层	卷积层	下采样层 * 2	—	卷积层 全连接层 softmax 层
D	10	484×484	卷积层	卷积层	下采样层 * 2	卷积层	卷积层	下采样层 * 2	卷积层	卷积层 全连接层 softmax 层
E	7	484×968	卷积层	下采样层 * 2	—	—	下采样层 * 2	上采样层 * 2	—	上采样层 * 2 卷积层 全连接层 softmax 层
F	8	484×968	卷积层	下采样层 * 2	卷积层	—	下采样层 * 2	上采样层 * 2	—	上采样层 * 2 卷积层 全连接层 softmax 层
G	9	484×968	卷积层	下采样层 * 2	卷积层	—	下采样层 * 2	上采样层 * 2	卷积层	上采样层 * 2 卷积层 全连接层 softmax 层
H	10	484×968	卷积层	下采样层 * 2	卷积层	卷积层	下采样层 * 2	上采样层 * 2	卷积层	上采样层 * 2 卷积层 全连接层 softmax 层

注:“—”表示该层为空。卷积层为:conv(kernel = 3×3 , stride = $\langle 1, 1 \rangle$, filter = 1);下采样层 * 2 为:conv(kernel = 3×3 , stride = $\langle 2, 2 \rangle$, filter = $n \times 2$)。上采样层 * 2;deconv(kernel = 3×3 , stride = $\langle 2, 2 \rangle$, filter = $n/2$);全连接层:dense(input, unit = 1 024, activation = ReLU)。

由于特征拼接矩阵同时包含量化特征矩阵和小鼠路径图像,故 E—H 网络结构相较 A—D 网络结构使用了更多的上采样和下采样操作,而 A—D 网络则以卷积结构为主。在小鼠运动的有限特征空间中,为了对小规模样本的期望进行近似估计,本文使用随机梯度下降法对样本集的损失函数进行最小化求解。训练阶段的分类损失 L 为

$$L = CE(Q, Q') \quad (8)$$

式中,数据集 Q 由 m 个特征矩阵 B_i 构成, Q' 代表对应 B_i 分类结果真值的标签集, $CE(\cdot)$ 代表交叉熵损失。

除了输出层以外,本文 A—H 网络的各层结构

如表 1 所示,各网络最后一层均使用 softmax 作为分类输出,即小鼠 OFT 视频的分类概率。为了防止网络过拟合,训练过程中令 dropout = 0.7,即在每次训练过程中,中间层只有 70% 的神经元进行工作。为了保证预测的准确率,在测试过程中所有的神经元均参与分类运算。

3 实验及结果分析

3.1 旷场实验数据及配置

为了观察实验小鼠注射了抗抑郁药物后对新

环境紧张的行为差异,旷场实验装置分为旷场反应箱和数据采集两部分,旷场反应测试箱是4个四壁黑色不透明的有机板箱(45 × 45 × 45 cm),箱底是白色有机板,中心为15 × 15 cm的正方形区域。在行为箱上放置一架分辨率为1 280 × 720像素的彩色摄像机记录小鼠的运动情况,如图3(a)所示。

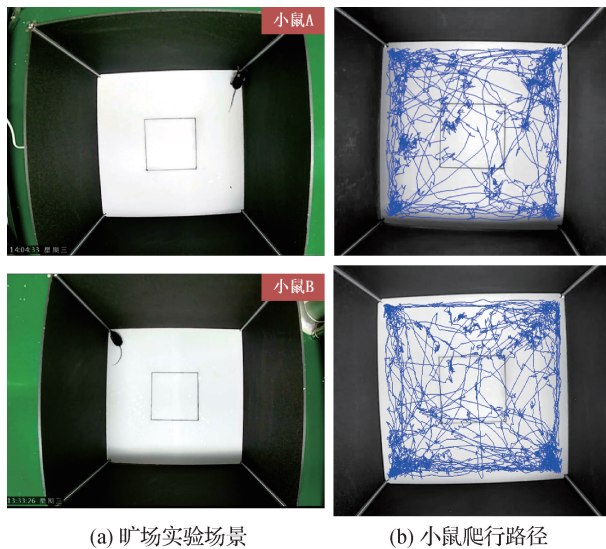


图3 旷场实验场景
Fig. 3 The OFT experiments setup((a)OFT experiments setups;(b)clawing path images)

实验开始前,将小鼠放在行为测试室适应30 min。实验时将小鼠背朝实验操作者轻柔地放入测试箱的正中央,让小鼠自由活动30 min,其数据不参与随后的计算和统计。由于小鼠在24 h内有其活动的周期,每次实验选择在相同时间段完成。实验同时拍摄32只小鼠的活动数据,其中17只测试组小鼠,15只参考组小鼠。测试组小鼠注射了某种抗抑郁试验药物,参考组均为正常未注射任何药物的小鼠。数据采集完成后,采用2.1节所述的OFT视频数据处理方法对可量化及非可量化特征进行提取,如图3和表2所示。图3中分别展示了小鼠A和小鼠B的拍摄模式及爬行路径图,表2为两只小鼠行为特征的定量计算结果。本文计算平台为Ubuntu17.04操作系统,采用Python语言在Intel(R) Core(TM) i7-6700 CPU@3.40 GHz、内存8 GB、GPU为Nvidia 2080Ti的双计算显卡上进行TensorFlow网络运算。

对旷场实验中拍摄的小鼠视频,首先进行特征

表2 小鼠爬行可量化特征计算结果
Table 2 Calculated results of quantifiable characteristics of mouse crawling

可量化特征	图3 小鼠 A	图3 小鼠 B
平均速度/(cm/s)	52.376 7	35.555 8
总距离/cm	5161.42	6992.34
休息时间/s	52.124	316.23
左上时间/s	112.88	131.306
右上时间/s	144.42	214.97
左下时间/s	118.856	208.994
右下时间/s	53.784	368.354
左边时间/s	112.88	131.306
右边时间/s	303.282	761.276
上边时间/s	329.51	621.67
下边时间/s	209.492	735.048
中间时间/s	43.824	51.958

提取与描述。为了使图像大小保持一致,统一将小鼠路径图像缩放到484 × 484像素,并归一化到-1.0~1.0之间以便深度学习网络训练和收敛。本文整体数据集集中有正样本3 080个,负样本1 034个,每个量化特征矩阵为484 × 484维,拼接后的特征矩阵为484 × 968维。在每次训练中分别等量抽取正负样本的75%用于训练网络模型,其余作为测试集数据。

对提取的小鼠特征矩阵B中的特征做框箱图,如图4所示,图中蓝色框箱为参考组视频特征,红色框箱为测试组小鼠视频特征。对比每维特征可以看出小鼠在各个特征维上均具有一定的分类特性。为了进行实验对比,分别使用了人工分类、经典的SVM分类、卷积神经网络分类3种分类方法。

3.2 小鼠 OFT 视频的人工及 SVM 分类实验

在人工分类实验中,首先要求16位测试者观看标记好的20个小鼠视频并分析两组小鼠运动行为的差异。测试时,选择另外20个小鼠视频,并去除小鼠视频标记,要求测试者将20个视频分为两类,然后统计分类的准确率以及正确分类数占总数的比例。在SVM分类实验中,所有实验应用单类向量分类器(one-class support vector classifier,V-SVC)模型(Schölkopf等,2001),通过比较选择使用线性核函数,并利用交叉验证寻得最优参数,对上述OFT

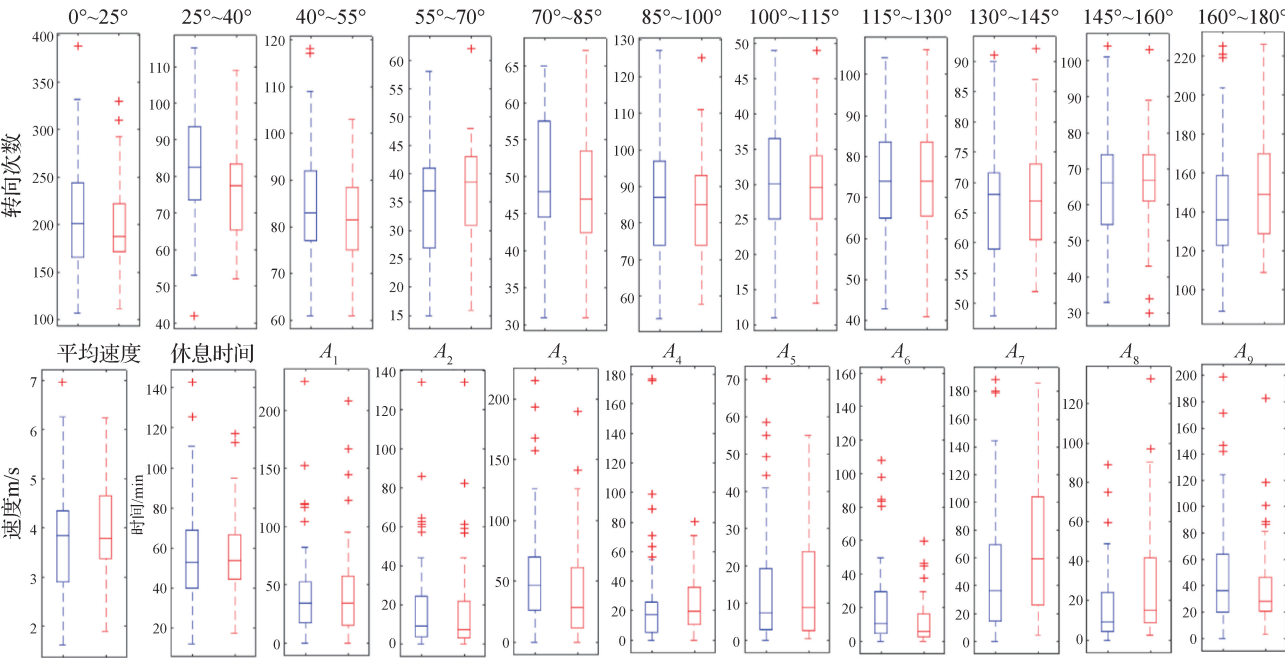


图 4 两组小鼠各个运动特征维度的量化对比

Fig.4 Quantitative comparisons of each dimension of mice behavioral features

视频的 22 维可量化特征进行二分类,得到最终的分
类精度为 75.19%。

为了进一步分析 SVM 分类中不同特征维度对
分类结果的影响,本文逐一将各个特征从 22 维特征
中去除,分别用所剩的 21 维进行 SVM 分类,其分类
准确率如表 3 所示。表中 A_1 — A_9 表示小鼠在 9 个
不同区域的停留时间,排名依据去除某一特征后对
准确率影响的大小,排名靠前代表影响显著。

通过分析可以发现,实验中小鼠在各个区域的
停留时间、速度和休息时间对分类的影响较为显著,
这些维度在表中的 Rank 排名居前 10 位,而运动过
程中的小角度转向频次对小鼠行为分类的影响较
小,值得注意的是,去除 $115^\circ \sim 130^\circ$ 或 $130^\circ \sim 145^\circ$
转向频次之后进行分类,结果的精度显著降低至
 38.09% 以下,说明这个区间的大角度运动转向频
次对分类起着关键作用。进而,采用特征拼接矩阵
进行 SVM 分类,使用 3 000 个特征拼接矩阵作为训
练样本,1 114 个样本作为测试样本,分类结果精度
为 80.07% 。与之前的 22 维特征相比结果,采用特
征拼接矩阵作为输入的分类精度提高了近 5% ,两
种不同输入的 SVM 分类精度参数如表 4 所示。

3.3 小鼠 OFT 视频的卷积神经网络分类

在 OFT 视频集中随机选取 $3/4$ 作为卷积神经

表 3 逐一去除某个特征维度后的 SVM 分类精度

Table 3 SVM classification precision after
removing a specific feature dimension

去除的特征维度	剩余特征分类准确率/%	排名
平均速度	68.18	16
休息时间	64.36	11
A_1 停留时间	66.66	14
A_2 停留时间	61.90	9
A_3 停留时间	57.14	6
A_4 停留时间	47.60	4
A_5 停留时间	59.42	7
A_6 停留时间	42.85	3
A_7 停留时间	53.33	5
A_8 停留时间	61.90	9
A_9 停留时间	59.42	7
$0^\circ \sim 25^\circ$	71.42	17
$25^\circ \sim 40^\circ$	73.33	19
$40^\circ \sim 55^\circ$	71.42	17
$55^\circ \sim 70^\circ$	66.66	14
$70^\circ \sim 85^\circ$	75.19	21
$85^\circ \sim 100^\circ$	75.19	21
$100^\circ \sim 115^\circ$	73.33	19
$115^\circ \sim 130^\circ$	38.09	2
$130^\circ \sim 145^\circ$	33.33	1
$145^\circ \sim 160^\circ$	64.36	11
$160^\circ \sim 180^\circ$	64.36	11

表 4 采用 22 维可量化特征与特征拼接矩阵作为输入的 SVM 分类比较

Table 4 SVM classification comparison between input with 22 dimensional quantitative features and spliced feature matrix

输入	支持向量数	标准支持向量数	边界上支持向量数	精度/%
22 维特征	63.00	35.00	28.00	75.19
484 维拼接特征	1 509.00	750.00	759.00	80.07

注:加粗字体为最优结果。

网络的训练数据集,剩余 1/4 的小鼠视频作为测试数据集。图 5 是 4 种不同网络结构 A—D 分别采用两种不同输入时分类精度随迭代次数的变化趋势。图 5(a)是量化特征矩阵作为输入的分类准确率随样本数的变化曲线,图 5(b)是量化特征矩阵作为输入的分类准确率随样本数的变化曲线。表 5 展示了采用小鼠路径图像及 484 维量化特征矩阵作为输入时不同网络结构的分类精度。由图 5 可以看出,分类准确率随着样本数的增加而逐渐趋于稳定,此时网络更深且更复杂的 C 和 D 网络未获得更优的分类结果,而当网络深度以及复杂度到达 A 网络深度时,分类精度达到测试中的最高值。对于分类准确率最高的 A 和 B 网络,采用 OFT 量化特征可以有效提高分类精度 4.63%。分析表 5 可知,与采用小鼠路径图像作为输入相比,采用量化特征作为

表 5 A—D 网络结构的分类精度

Table 5 The precision of classification using A—D network structures

网络模型	小鼠路径图像	484 × 484 维特征矩阵
		/%
A 网络	91.66	96.29
B 网络	91.66	96.29
C 网络	82.75	92.59
D 网络	82.75	92.59

注:加粗字体为每行最优结果。

输入可以显著提高 OFT 视频的分类准确率,这也证明了本文采用量化特征矩阵作为 OFT 视频分类输入的合理性。

图 6 进一步对比了人工分类方法、SVM 分类方法与本文神经网络分类方法的准确率及其随样本数量变化的差异。可以看出,随着样本量的增加,除了人工分类方法外,其余方法的分类效率均在显著提升。而人工分类精度总体在 0.5 左右波动,表明测试人员仅根据人工观察很难对小鼠视频进行分类和判断,而随着样本量的增加,提高样本量并未有效提升人工分类准确率,从侧面反映出 OFT 视频分类对人工经验的要求较高,经过短期训练的测试者仍不足以高质量地完成 OFT 视频的准确分类任务。然而,随着样本量的增加,SVM 方法与本文分类方法的准确率均有所提升,本文采用的神经网络分类方

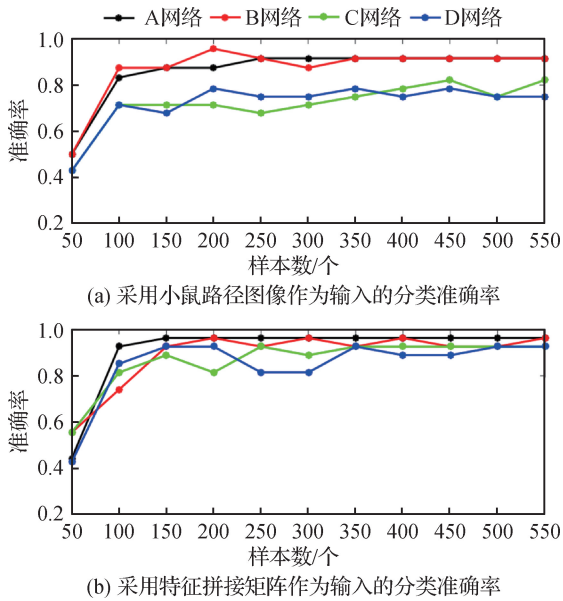


图 5 小鼠视频的分类准确率随样本数变化曲线
Fig. 5 The curves of classification precision with the number of OFT video samples((a)precision of classification using only the input of crawling path images of mice;(b)precision of classification using the input of the proposed spliced feature matrices)

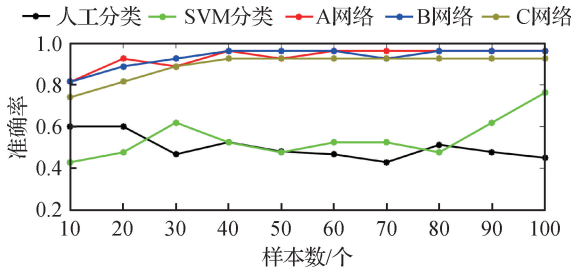


图 6 不同分类方法分类准确率比较及随样本数变化趋势
Fig. 6 Comparison of classification precision with different classification methods as increasing the number of samples

法显著优于 SVM 分类方法,表明采用本文深度神经网络对 OFT 视频分类的有效性。

图 7 绘制了采用特征拼接矩阵作为输入的 4 种网络结构的收敛曲线和分类精度曲线。图 7(a) 为损失值随迭代次数的变化曲线;图 7(b) 为分类精度随迭代次数的变化曲线。由图 7 可见, G 网络结构能够获得较高的分类性能,其收敛程度与分类精度均优于 E 和 H 网络,略优于 F 网络。其中 H 网络收敛过程中的较大抖动,反映出采用更深参数量更大的网络结构容易在小样本条件下形成过拟合状态,从而导致网络整体收敛于局部最优的情况。4 种不同网络结构 E—H 的分类精度与损失值的数值比较由表 6 给出。

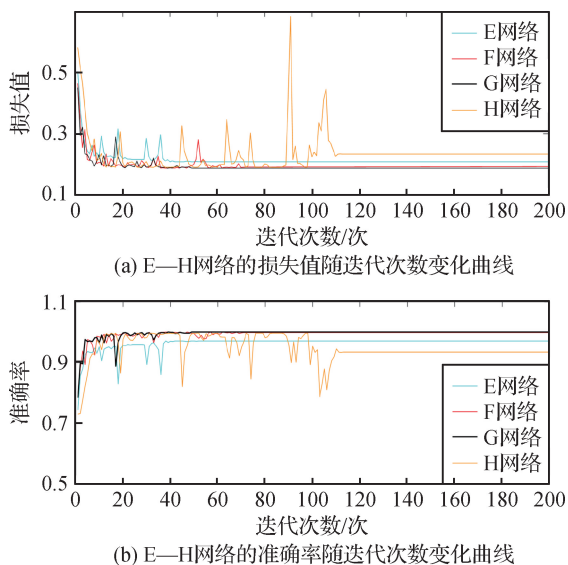


图 7 E—H 网络对 OFT 视频分类的损失值与准确率曲线

Fig. 7 The loss values and precision curves of the OFT video classification based on E—H networks((a) curves of the loss values of E—H network with the number of iterations;(b) curves of the classification precision of E—H network with the number of iterations)

表 6 E—H 网络对 OFT 视频的分类准确率与损失值比较
Table 6 Comparison of precision and training loss of OFT video classification based on E—H network

网络名称	损失值	分类准确率/%
E 网络	0.207 4	97.08
F 网络	0.192 3	98.83
G 网络	0.187 3	99.25
H 网络	0.230 7	93.39

注:加粗字体为各列最优结果。

此外,为了比较训练集与测试集样本比例变化对收敛及分类准确率的影响,采用 G 网络进行不同比例的测试,结果如图 8 所示。其中不同颜色曲线表示测试集占总样本的比例,上半部分曲线代表准确率曲线,下半部分代表损失值收敛曲线。由图 8 可见,随着测试集占比的增大,分类准确率逐步降低。测试集占比小于 0.3 时,分类准确率降低并不明显。测试集占比为 0.5~0.9 时,分类准确率明显下降。测试集占比达到 0.9 时,分类准确率已降至 0.7 以下(为了与下半部分曲线显示区分,本文对 0.7 以下的准确率数值进行了截断处理)。依据上述实验结果,通常情况下单次测试数据集与训练数据集比例为 0.25 时分类结果的准确率较优。

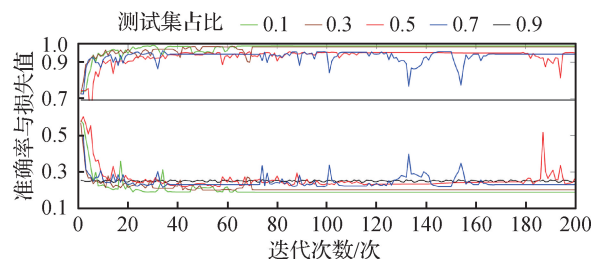


图 8 G 型结构下不同测试数据集占比对分类准确率及损失值的影响对比

Fig. 8 Comparison of the influence of different proportions of test datasets on classification precision and training loss based on G network

现有视频分类算法大多将 RGB 视频帧与运动光流作为输入,侧重在标准视频分类数据集如 Hollywood2、Columbia Consumer Video、Kinetics、Something-something、UCF101 和 HMDB 等上进行基于内容的分类。考虑到本文 OFT 视频分类任务的特殊性,与 Jiang 等人(2018)提出动态递归的神经网络(recurrent dynamic neural network, rDNN)、类间关系的视频分类方法和零样本学习(zero-shot learning, ZSL)网络(Brattoli 等,2020)进行实验对比。采用两种网络的预训练模型的分类结果准确率如表 7 所示。可以看出,两种主流深度学习视频分类方法未实现准确分类。导致这种结果的原因或可归结为两点:1)两类方法主要考虑视频内容上的差异或运动目标短时的运动特征,而 OFT 视频分类侧重对小鼠运动特征的长时统计差异,故本文方法的特征提取在计算长时的统计差异方面具有优势。2)rDNN 网络等侧重对单人或多人的行为进行分类与识别,其

多分类问题的类间差异较为明显,而 OFT 视频的类间差异显著小于上述公开数据集的类间差异。

在计算复杂性方面,以 G 型网络为例,本文方法的网络参数量为 1.2 M,计算量为 4.7 Mflops,预测每段视频时间约 1.6 s 左右。为了统一对比标准,在计算效率评估阶段,本文选取了 200 段时长为 1.5 min 的 OFT 视频,计算各种方法每段视频处理时间的平均值,结果见表 7。可以看出,本文方法量化特征与非量化特征提取部分耗时较长,故总体的运行效率不及 rDNN 网络和 SVM 方法。

表 7 不同视频分类网络对 OFT 视频的
分类准确率与运行时间对比

Table 7 Comparison of precision on OFT
video classification with different networks

网络	分类准确率/%	运行时间/s
SVM	80.07	19.1
rDNN	62.5	9.7
ZSL	74.64	461.4
G 网络(本文)	99.25	87.9

注:加粗字体为各列最优结果。

4 结 论

针对 OFT 视频对小鼠运动行为进行差异比问题,本文提出一种基于特征拼接矩阵的卷积神经网络分类方法。首先,通过对小鼠视频中爬行路径、速度、转弯和停留等运动行为特征的计算量化,构建小鼠运动特征的可量化特征矩阵。其次,结合非量化的小鼠爬行路径图像,提出采用特征拼接矩阵方式对可量化特征与非量化特征进行统一描述。在此基础上,采用不同的卷积神经网络对特征拼接矩阵进行学习和分类。

实验基于真实 OFT 实验视频数据集,实验组小鼠和参考组小鼠分别注射某种抗抑郁药物和安慰剂。通过对 OFT 视频片段的学习,本文所提基于深度学习的小鼠视频分类方法取得了比人工分类及传统 SVM 方法更优的分类结果。本文工作证明采用层数不多的简单深度神经网络结构即可以获得准确的 OFT 视频分类结果。在实验数据集上,本文方法 OFT 视频分类的精度为 99.25%,有效降低了 OFT

数据分析人员的工作量,提高了对大规模 OFT 视频分类的准确率。此外由实验结果发现,小鼠的停留区域与时间、大角度转向频次等特征对分辨小鼠的重要性高于其他维度信息。

受 OFT 实验场景以及数据量的限制,仅在较为单一的 OFT 实验数据集上测试了提出的方法,未测试在不同类型 OFT 实验视频数据集上的性能,故本文方法的泛化性仍有待验证。未来工作将进一步研究小鼠可分辨特征与实际行为的关系,并探索更具泛化性的基于深度学习的 OFT 视频分类方法。

致 谢 感谢西安交通大学第一附属医院转化医学中心朱峰教授前期提供小鼠旷场实验视频与极具建设性的讨论。感谢西北工业大学周果清副教授在 SVM 以及深度学习算法实现方面的讨论和帮助。

参考文献(References)

Brattoli B, Tighe J, Zhdanov F, Perona P and Chalupka K. 2020. Rethinking zero-shot video classification: end-to-end training for realistic applications//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 4612-4622 [DOI: 10.1109/CVPR42600.2020.00467]

Burgos-Artiz X P, Dollár P, Lin D Y, Anderson D J and Perona P. 2012. Social behavior recognition in continuous video//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA; IEEE: 1322-1329 [DOI: 10.1109/CVPR.2012.6247817]

Carreira J and Zisserman A. 2017. Quo vadis, action recognition? A new model and the kinetics dataset//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 4724-4733 [DOI: 10.1109/CVPR.2017.502]

Choleris E, Thomas A W, Kavaliers M and Prato F S. 2001. A detailed ethological analysis of the mouse open field test: effects of diazepam, chlordiazepoxide and an extremely low frequency pulsed magnetic field. Neuroscience and Biobehavioral Reviews, 25(3): 235-260 [DOI: 10.1016/S0149-7634(01)00011-2]

Cortes C and Vapnik V. 1995. Support-vector networks. Machine Learning, 20(3): 273-297 [DOI: 10.1007/BF00994018]

Decker C M and Hamprecht F A. 2014. Detecting individual body parts improves mouse behavior classification//Proceedings of the Workshop on Visual Observation and Analysis of Vertebrate and Insect Behavior, Stockholm, Sweden; #2089814

Deng L and Yu D. 2014. Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3/4): 197-387 [DOI: 10.1561/20000000039]

Durand T, Thome N and Cord M. 2015. MANTRA: minimum maximum

- latent structural SVM for image classification and ranking//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE; 2713-2721 [DOI: 10.1109/ICCV.2015.311]
- Feng W G, Gao J, Buckles B P and Wu K W. 2011. Wireless capsule endoscopy video classification using an unsupervised learning approach. *Journal of Image and Graphics*, 16(11): 2041-2046 (冯文刚, 高隽, Buckles B P, 吴克伟. 2011. 无监督学习的无线胶囊内镜视频分类方法. *中国图象图形学报*, 16(11): 2041-2046) [DOI: 10.11834/jig.20111113]
- Freund Y and Schapire R E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1): 119-139 [DOI: 10.1006/jcss.1997.1504]
- Giancardo L, Sona D, Huang H P, Sannino S, Managò F, Scheggia D, Papaleo F and Murino V. 2013. Automatic visual tracking and social behaviour analysis with multiple mice. *PLoS One*, 8(9): e74557 [DOI: 10.1371/journal.pone.0074557]
- Gu J X, Ding X Q and Wang S J. 2009. A survey of activity analysis algorithms. *Journal of Image and Graphics*, 14(3): 377-387 (谷军霞, 丁晓青, 王生进. 2009. 行为分析算法综述. *中国图象图形学报*, 14(3): 377-387) [DOI: 10.11834/jig.20090301]
- Hall C S. 1934. Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *Journal of Comparative Psychology*, 18(3): 385-403 [DOI: 10.1037/h0071444]
- Hong W Z, Kennedy A, Burgos-Artizzu X P, Zelikowsky M, Navonne S G, Perona P and Anderson D J. 2015. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38): E5351-E5360 [DOI: 10.1073/pnas.1515982112]
- Jiang Y G, Wu Z X, Wang J, Xue X Y and Chang S F. 2018. Exploiting feature and class relationships in video categorization with regularized deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(2): 352-364 [DOI: 10.1109/TPAMI.2017.2670560]
- Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R and Li F F. 2014. Large-scale video classification with convolutional neural networks//Proceedings of 2014 IEEE conference on Computer Vision and Pattern Recognition. Columbus, USA; IEEE; 1725-1732 [DOI: 10.1109/CVPR.2014.223]
- Katsageorgiou V M, Zanotto M, Huang H P, Ferretti V, Papaleo F, Sona D and Murino V. 2016. Unsupervised mouse behavior analysis: a data-driven study of mice interactions//Proceedings of the 23rd International Conference on Pattern Recognition (ICPR). Cancun, Mexico; IEEE; 925-930 [DOI: 10.1109/ICPR.2016.7899754]
- Lin W H and Hauptmann A. 2002. News video classification using SVM-based multimodal classifiers and combination strategies//Proceedings of the 10th ACM International Conference on Multimedia. Juan-les-Pins, France; ACM; 323-326 [DOI: 10.1145/641007.641075]
- Lu B, Wei X K and Bi D Y. 2005. Applications of support vector machines in classification. *Journal of Image and Graphics*, 10(8): 1029-1035 (陆波, 尉询楷, 毕笃彦. 2005. 支持向量机在分类中的应用. *中国图象图形学报*, 10(8): 1029-1035) [DOI: 10.11834/jig.200508190]
- Monteiro J P, Oliveira H P, Aguiar P and Cardoso J S. 2014. A depth-map approach for automatic mice behavior recognition//Proceedings of 2014 IEEE International Conference on Image Processing (ICIP). Paris, France; IEEE; 2261-2265 [DOI: 10.1109/ICIP.2014.7025458]
- Nair A, Srinivasan P, Blackwell S, Alceick C, Fearon R, De Maria A, Panneershelvam V, Suleyman M, Beattie C, Petersen S, Legg S, Mnih V, Kavukcuoglu K and Silver D. 2015. Massively parallel methods for deep reinforcement learning [EB/OL]. <https://arxiv.org/pdf/1507.04296.pdf>
- Nam H and Han B. 2016. Learning multi-domain convolutional neural networks for visual tracking//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE; 4293-4302 [DOI: 10.1109/CVPR.2016.465]
- Ng J Y H, Hausknecht M, Vijayanarasimhan S, Vinyals O, Monga R and Toderici G. 2015. Beyond short snippets: deep networks for video classification//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE; 4694-4702 [DOI: 10.1109/CVPR.2015.7299101]
- Nováková J, Vinklerová J, Kellnerová R and Šulcová A. 1997. Computer analysis of video-taped agonistic behaviour in mice: effects of interferon- α , diazepam and naloxone. *Scripta Medica*, 70(7): 267-274
- Schölkopf B, Platt J C, Shawe-Taylor J, Smola A J and Williamson R C. 2001. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7): 1443-1471 [doi: 10.1162/089976601750264965]
- Schuldts C, Laptev I and Caputo B. 2004. Recognizing human actions: a local SVM approach//Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, UK; IEEE; 32-36 [DOI: 10.1109/ICPR.2004.1334462]
- Stanford S C. 2007. The open field test: reinventing the wheel. *Journal of Psychopharmacology*, 21(2): 134-135 [DOI: 10.1177/0269881107073199]
- Svetnik V, Liaw A, Tong C, Culberson J C, Sheridan R P and Feuston B P. 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, 43(6): 1947-1958 [DOI: 10.1021/ci034160g]
- Walsh R N and Cummins R A. 1976. The open-field test: a critical review. *Psychological Bulletin*, 83(3): 482-504 [DOI: 10.1037/

0033-2909. 83. 3. 482]

- Wang J, Yang Y, Mao J H, Huang Z H, Huang C and Xu W. 2016. CNN-RNN: a unified framework for multi-label image classification//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2285-2294 [DOI: 10.1109/CVPR.2016.251]
- Xu Z Q, Hu J N and Deng W H. 2016. Recurrent convolutional neural network for video classification// Proceedings of 2016 IEEE International Conference on Multimedia and Expo (ICME). Seattle, USA: IEEE: 1-6 [DOI: 10.1109/ICME.2016.7552971]
- Zhou W S, Vellaikal A and Kuo C C J. 2000. Rule-based video classification system for basketball video indexing//Proceedings of 2000 ACM Workshops on Multimedia. Los Angeles, California, USA: ACM: 213-216 [DOI: 10.114 5/357744.357941]

作者简介



肖照林, 1984 年生, 男, 副教授, 主要研究方向为计算摄像和计算机视觉。

E-mail: xiaozhaolin@xaut.edu.cn



金海燕, 通信作者, 女, 教授, 主要研究方向为计算机视觉、图像处理、智能信息处理。

E-mail: jinhaiyan@xaut.edu.cn

杨志林, 男, 硕士研究生, 主要研究方向为图像处理和计算机视觉。E-mail: zlin_yang@126.com

刘欢, 男, 硕士研究生, 主要研究方向为计算摄像和计算机视觉。E-mail: liuhuanjsj@gmail.com