计算机软件新技术国家重点实验室（南京大学）

**State Key Laboratory of Novel Software Technology (Nanjing University)**

**巫义锐老师:**

您好！按照南京大学计算机软件新技术国家重点实验室开放课题基金管理细则，课题结束后，需向实验室提交如下资料：

(1) 研究工作总结。

(2) 学术论文与/或其他成果证明。

发表的论文、著作请写明：题目、作者(包括单位)、刊物卷期(出版时间，出版社)或会议名称、页码等。

**请于2019年7月15日前E-mail发给我室，谢谢合作（已提前提交结题报告的无需重复提交）。**

**南京大学计算机软件新技术国家重点实验室**

2019年6月3日

附：

1、开放基金资助项目总结报告填写完后E-mail至: [keysoftlab@nju.edu.cn](mailto:keysoftlab@nju.edu.cn)。

2、论文有正式电子版的，发送电子版即可；如无电子版，请将全文复印件等材料寄到我处，地址：

210023 江苏省南京市栖霞区仙林大道163号南京大学仙林校区机关603信箱计算机系 徐烜（18912990005） 收。

**南京大学计算机软件新技术国家重点实验室**

**开放基金资助项目总结报告**

|  |  |  |  |
| --- | --- | --- | --- |
| 项目批准号 | KFKT2017B05 | 项目名称 | 基于文字对称性与上下文关系的自然场景文字检测研究 |
| 项目负责人 | 巫义锐 | 所在单位 | 河海大学计算机与信息学院 |
| **基于文字对称性与上下文关系的自然场景文字检测研究学术总结**  **1研究结果：**  本项目瞄准文字检测在自然场景理解领域高速发展与广泛应用的契机，针对已提出方法所忽略的利用“**文字对称性**”、利用“**场景上下文信息**”、和提高“**文字检测效率**”的问题，基于场景文字的固有特性，提出了文字对称特性建模，场景上下文信息建模与基于形状特征的特征建模，提出了准确、鲁棒和高效的自然场景文字检测算法。  具体来说，ICDAR17的文章《A Robust Symmetry-based Method for Scene/Video Text Detection Through Neural Network》构建了笔画像素值的类间对称（笔画像素对的梯度与灰度值相似）和笔画间的同类对称（同一字符笔画的恒定宽度和笔画的规整颜色、纹理分布），并将其用于加强特征描述子对文字区域与非文字区域辨别能力，最终在文字识别标准数据集ICDAR2015数据集上上取得了0.72的准确率，0.39的召回率。与之相对应的，NJU Text实现了0.70的准确率，0.36的召回率。SCI期刊Computational Intelligence的文章《Deep spatio-temporal LSTM network with temporal pattern feature for 3D human action recognition》利用LSTM模型直接对多尺度场景上下文信息进行编码，提升了基于骨骼的动作识别精度。PCM17的文章《Cloud of Line Distribution for Arbitrary Text Detection in Scene/Video/License Plate Images》通过计算关键点间的角度与梯度信息描述文字候选区域的形状，进而推断文字候选区域是否为文字，最终在车牌识别数据集Medialab LPR上实现了0.90的准确率，0.81的召回率与1.23s每一幅图像的运算效率。与之对应的，Zambeerletti在15年ACPR所提算法只能达到0.83的准确率，0.76的召回率与0.61s每一幅图像的运算效率。  由此可见，所做工作良好的反映了所写计划书中的要点，从多个角度提高自然场景文字的检测率、运算效率以及针对不同特性文字的识别鲁棒性，进而一定程度上消除了场景文字与文档文字间的识别鸿沟。  **2研究结果的科学意义及学术界引用：**  本项目围绕基于对称性与上下文信息的自然场景文字检测研究这一科学问题，通过整合多类型异构特征与场景文字上下文信息，从文字候选区域生成、单字符检测、文字行生成、实时算法性能等多个方面分析场景文字检测，为场景文字检测这一重要的计算机视觉领域的发展做出有特色的贡献。本项目的创新之处主要体现在： （1）通过对文字本征特性-对称性的多领域描述以及自适应特征融合，提升了文字检测算法针对于不同文字检测挑战的鲁棒性。研究文字对称性在多领域内的描述模型，将对称特征与外观特征看作异构特征，并进行自适应融合。该研究对提升文字检测鲁棒性和准确性提供方法创新。 （2）构建场景上下文信息为依据的文字行检测问题。研究场景上下文信息建模方法，并将其应用于骨骼动作识别算法中，提升了骨骼动作识别的鲁棒性与准确性。该研究为合理构建与利用上下文信息提供了方法创新。 （3）通过文字检测算法的设计与优化，实现了实时和准确的文字检测。通过低复杂度算法的设计，对文字区域的形状信息建模，实现了实时文字检测。该研究强调了文字检测效率问题，为依据应用环境，合理设计和优化文字检测算法提供了方法创新。  所发文章被多次引用。例如英国帝国理工学院的Guang-Zhong Yang（英国皇家工程院院士、 IEEE Fellow)教授在IEEE Trans. Medical Imaging发表的论文中多次引用、评价我们的工作《[End-to-end chromosome Karyotyping with data augmentation using GAN](javascript:void(0))》 “形成了自动智能的分类方法”与“对不充分数据实现了基于深度智能方法的数据增强”。 Le Yan等人在IMCOM2019的文章《Small Watershed Stream-Flow Forecasting Based on LSTM》论文中引用我们的工作《 Local and global bayesian network based model for flood prediction》，并对通过LSTM模型上下文建模方法给予了较高的评价。  **3.存在的不足**  1. 由于首先发表了一些会议文章，然后将会议文章改为期刊文章，导致大量文章仍然在审稿过程中，整体文章所发表会议或期刊的层次较低。  2. 基于上下文建模的场景文字识别方法难度较高，未能被所负责的研究生同学完成任务，近来已经被另外一位同学接手，希望能在较短时间内形成结果，补足整个体系中的  一些不足。  3. 所发表文章聚焦度不够，涵盖了较多的相关课题，希望能在以后的项目实现中更加聚焦于小的子领域。  项目负责人： 无标题.jpg  2019年 7 月 14 日 | | | |

|  |  |
| --- | --- |
|  | 结合研究兴趣与研究基础，我们提出了“基于对称性与上下文信息的自然场景文字检测研究”的课题。该课题一方面继承了本项目组在自然场景文字检测与识别的研究经验，另一方面又具有前瞻性、创新性与实用性。本项目组成员对项目的研究目标、研究内容及具体研究方案进行了反复深入的讨论，形成了缜密、科学、合理的研究思路和方案，并将其付诸实践，在一些关键问题上形成了论文。  在理论与方法研究方面，本项目组对三个关键研究内容进行了探索，验证了实验方案。首先，对于文字对称特性建模及特征融合方法做了一些工作，其中文字对称特性建模的相关研究成果已发表在ICDAR17论文《A Robust Symmetry-based Method for Scene/Video Text Detection Through Neural Network》上，特征融合方法的相关研究结果发表在MMM18论文《A Novel 3D Human Action Recognition Framework for Video Content Analysis》上。其次，场上下文信息建模算法研究也有一定的成果，具体为LSTM模型的建模以及注意力机制的建模，发表在Computational Intelligence国际期刊上，及ICPR18《Context-Aware Attention LSTM Network for Flood Prediction》等国际重要期刊与学术会议上。此外，本项目组针对实时场景文字检测方法也着手进行研究，其中部分算法复杂度优化的相关工作已发表于ACPR，PCM等国际重要期刊与学术会议上。以上前期工作有效地证明了所提出方法的合理性。最终，项目组总计以此项目为依托，发表论文篇，其中EI论文篇，SCI论文两篇。  项目组建立并完善的场景文字检测系统，项目组受理4项相关中国发明专利与软件著作。基于系统与专利，完成了与杭州全速科技网络公司合作，实现了相关人工智能技术的转化，实现了技术研究收入10万元，并不断的在深化合作中。  项目组成员将积极参加模式识别、计算机视觉和多媒体等领域的重要学术会议，共参加了如ICDAR、ICPR、ICIP、PCM、MMM等，加强与国内外相关研究方向专家和学者的交流。另一方面，项目组除进一步加强与南京大学、印度统计研究所、马来亚大学的交流互访外，还积极邀请国内外知名专家到项目组依托单位访问交流，争取建立多渠道的合作与交流。 |
|  |
| 研 |
|  |
| 究 |
|  |
| 工 |
|  |
| 作 |
|  |
| 总 |
|  |
| 结 |
|  |
| 摘 |
|  |
| 要 |
|  |
|  |
|  |
| 主  题  词 | 文字对称特性建模，深度学习，注意力机制，特征融合，实时系统 |

**南京大学计算机软件新技术国家重点实验室**

**开放基金资助项目结题简表**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 批准号 | | | KFKT2017B05 | | | | | | | 项目名称 | | | | | | | 基于文字对称性与上下文关系的自然场景文字检测研究 | | | | | | | | | |
| 项目负责人 | | | 巫义锐 | | | | | | | 专业技术职务 | | | | | | | 副教授 | | | | | | | | | |
| 所在单位 | | | 河海大学 | | | | | | | 单位性质 | | | | | | | ☑高等院校 B科研单位  C其它 | | | | | | | | | |
| 实际参加人数 | | | 高级 | | | 中级 | | | | | 初级 | | | 博士后 | | | | | | 博士生 | | | 硕士生 | | | |
|  | | | 1 | | |  | | | | |  | | |  | | | | | | 1 | | | 4 | | | |
| 研究经费 | | | 资助总额 | | | 2.5 | | | | | 研究期限 | | | 计划完成年月 | | | | | | | | 2019-6-30 | | | | |
| 实际完成年月 | | | | | | | | 2019-6-30 | | | | |
| 完 | 学术会议交流 | | | | | | | | 专著出版 | | | | | | |  | | 期刊发表 | | | | | | | | |
| 成  论 |  | 大会特  邀报告 | | | 分组报告  或展示 | | | | 已出版 | | | | 拟出版 | | |  | | 全国性科技  期刊 | | | | | | 国外学  术刊物 | | |
| 著  (篇 | 国际 |  | | | 5 | | | |  | | | |  | | |  | | 已发表 | | | 待发表 | | | 已发表 | | 待发表 |
| 册) | 国内 |  | | | 6 | | | |  | | | |  | | |  | |  | | |  | | | 3 | | 1 |
| 结  题 | 论著评审 | | | 审鉴定 | | | | 会议审定 | | | | | 发明专利 | | | | | | 评审鉴定日期 | | | | | | | |
| 形  式 | √ | | |  | | | |  | | | | |  | | | | | | 2019.7.15 | | | | | | | |
| 人  才 | 博士后 | | | | | | | 博士 | | | | | | | | | | | 硕士 | | | | | | | |
| 培  养 |  | | | | | | |  | | | | | | | | | | | 1 | | | | | | | |
| 学 | 国家级 | | | | | | | | | | | 省部级 | | | | | | | | | 国际（学术〕奖 | | | | 其它 | |
| 术  奖 | 自然科学奖 | | 科技进步奖 | | | | 发明奖 | | | | | 自然科学奖 | | | 科技进步奖 | | | | | |  | | | |  | |
| 励 |  | |  | | | |  | | | | |  | | |  | | | | | |

**南京大学计算机软件新技术国家重点实验室**

**开放基金资助项目完成论著目录**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 论著名称 | 作者 | 作者单位 | 发表时间 | 刊物\会议名称 | 卷、期、页码 | SCI、EI检索号码 | CCF等级 |
| **An Automatic System for Generating Artificial Fake Character Images** | [Yisheng Yue](https://dblp.uni-trier.de/pers/hd/y/Yue:Yisheng), [Palaiahnakote Shivakumara](https://dblp.uni-trier.de/pers/hd/s/Shivakumara:Palaiahnakote), Yirui Wu, [Liping Zhu](https://dblp.uni-trier.de/pers/hd/z/Zhu:Liping), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong), [Umapada Pal](https://dblp.uni-trier.de/pers/hd/p/Pal_0001:Umapada) | 南京大学，河海大学，马来亚大学 | 2019.1.8 | International Conference on Multimedia modeling | 291-301 | 20190306375679 | C |
| **Hierarchical Bayesian Network Based Incremental Model for Flood Prediction** | Yirui Wu, [Weigang Xu](https://dblp.uni-trier.de/pers/hd/x/Xu:Weigang), [Qinghan Yu](https://dblp.uni-trier.de/pers/hd/y/Yu:Qinghan), [Jun Feng](https://dblp.uni-trier.de/pers/hd/f/Feng:Jun), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong) | 南京大学，河海大学 | 2019.1.8 | International Conference on Multimedia modeling | 556-566 | 20190306375753 | C |
| **End-To-End Chromosome Karyotyping with Data Augmentation Using GAN** | Yirui Wu, [Yisheng Yue](https://dblp.uni-trier.de/pers/hd/y/Yue:Yisheng), [Xiao Tan](https://dblp.uni-trier.de/pers/hd/t/Tan:Xiao), [Wei Wang](https://dblp.uni-trier.de/pers/hd/w/Wang:Wei), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong): | 南京大学，河海大学， | 2018.10.7 | International Conference on Image Processing | 2456-2460 | 20191206646613 | C |
| **Local and Global Bayesian Network based Model for Flood Prediction** | Yirui Wu, [Weigang Xu](https://dblp.uni-trier.de/pers/hd/x/Xu:Weigang), [Jun Feng](https://dblp.uni-trier.de/pers/hd/f/Feng:Jun), [Shivakumara Palaiahnakote](https://dblp.uni-trier.de/pers/hd/p/Palaiahnakote:Shivakumara), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong): | 南京大学，河海大学，马来亚大学 | 2018.8.20 | International Conference on Pattern Recognition | 225-230 | 20190206369074 | C |
| **Context-Aware Attention LSTM Network for Flood Prediction** | Yirui Wu, [Zhaoyang Liu](https://dblp.uni-trier.de/pers/hd/l/Liu:Zhaoyang), [Weigang Xu](https://dblp.uni-trier.de/pers/hd/x/Xu:Weigang), [Jun Feng](https://dblp.uni-trier.de/pers/hd/f/Feng:Jun), [Shivakumara Palaiahnakote](https://dblp.uni-trier.de/pers/hd/p/Palaiahnakote:Shivakumara), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong): | 南京大学，河海大学，马来亚大学 | 2018.8.20 | International Conference on Pattern Recognition | * 1301-1306 | 20190206369135 | C |
| **Em-SLAM: a Fast and Robust Monocular SLAM Method for Embedded Systems** | Yirui Wu, [Zhikai Li](https://dblp.uni-trier.de/pers/hd/l/Li:Zhikai), [Shivakumara Palaiahnakote](https://dblp.uni-trier.de/pers/hd/p/Palaiahnakote:Shivakumara), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong): | 南京大学，河海大学，马来亚大学 | 2018.8.20 | International Conference on Pattern Recognition | 1882-1887 | 20190206369381 | C |
| **A Novel 3D Human Action Recognition Framework for Video Content Analysis** | [Lianglei Wei](https://dblp.uni-trier.de/pers/hd/w/Wei:Lianglei), Yirui Wu, [Wenhai Wang](https://dblp.uni-trier.de/pers/hd/w/Wang:Wenhai), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong): | 南京大学，河海大学 | 2018.2.5 | International Conference on Multimedia modeling | 48-60 | 20180804811760 | C |
| **Cloud of Line Distribution for Arbitrary Text Detection in Scene/Video/License Plate Images** | [Wenhai Wang](https://dblp.uni-trier.de/pers/hd/w/Wang:Wenhai), Yirui Wu, [Palaiahnakote Shivakumara](https://dblp.uni-trier.de/pers/hd/s/Shivakumara:Palaiahnakote), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong): | 南京大学，河海大学，马来亚大学 | 2017.9.28 | Pacific-Rim Conference on Multimedia | * 433-443 | 20180804812373 | C |
| **Cloud of Line Distribution and Random Forest Based Text Detection from Natural/Video Scene Images** | [Wenhai Wang](https://dblp.uni-trier.de/pers/hd/w/Wang:Wenhai), Yirui Wu, [Palaiahnakote Shivakumara](https://dblp.uni-trier.de/pers/hd/s/Shivakumara:Palaiahnakote), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong) | 南京大学，河海大学，马来亚大学 | 2018.2.5 | International Conference on Multimedia modeling | 48-60 | 20180804812373 | C |
| **Context and Temporal Aware Attention Model for Flood Prediction** | [Zhaoyang Liu](https://dblp.uni-trier.de/pers/hd/l/Liu:Zhaoyang), Yirui Wu, [Yukai Ding](https://dblp.uni-trier.de/pers/hd/d/Ding:Yukai), [Jun Feng](https://dblp.uni-trier.de/pers/hd/f/Feng:Jun), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong): | 南京大学，河海大学 | 2018.9.21 | Pacific-Rim Conference on Multimedia | 545-555 | 20184806162049 | C |
| **Compressing YOLO network by compressive sensing** | Yirui Wu, [Zhou-Yu Meng](https://dblp.uni-trier.de/pers/hd/m/Meng:Zhou=Yu), [Shivakumara Palaiahnakote](https://dblp.uni-trier.de/pers/hd/p/Palaiahnakote:Shivakumara), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong) | 南京大学，河海大学，马来亚大学 | 2017.12 | IAPR Asian Conference on Pattern Recognition | 19-24 | 20190506439402 |  |
| **Deep spatiotemporal LSTM network with temporal pattern feature for 3D human action recognition** | Yirui Wu，Lianglei Wei，Yucong Duan | 南京大学，河海大学，海南大学 | 2019.3.23 | Computational Intelligence | Online |  | C |
| **Channel-wise attention model-based fire and rating level detection in video** | Yirui Wu，Yuechao He，Palaiahnakote Shivakumara，Ziming Li，Hongxin Guo，Tong Lu | 南京大学，河海大学，马来亚大学 | 2019.4.21 | CAAI Trans. Intell. Techno | Online，1-5 |  |  |
| **Compressive Sensing based Convolutional Neural Network for Object Detection** | Yirui Wu, [Zhou-Yu Meng](https://dblp.uni-trier.de/pers/hd/m/Meng:Zhou=Yu), [Shivakumara Palaiahnakote](https://dblp.uni-trier.de/pers/hd/p/Palaiahnakote:Shivakumara), [Tong Lu](https://dblp.uni-trier.de/pers/hd/l/Lu:Tong) | 南京大学，河海大学，马来亚大学 | 2018.5 | Malaysian Journal of Computer Science | 待发表 |  |  |
| 基于网络爬虫的水利信息检索系统的设计与实现 | 巫义锐，黄多辉，周逸徉 | 南京大学，河海大学 | 2017.8 | 水 利 信 息 化 | 4 |  |  |

附论文全文