北京大学王选计算机研究所“2021本科生科研基金”课题一览

王选计算机研究所欢迎有志于计算机应用技术研究的理科院系本科生申报以下校长基金课题。

网上申请地址：<http://162.105.209.64/benyan>

申请截止日期：**2021年6月20日**

联系人：戴永宁 电话： 82529521 E-MAIL: daiyongning@pku.edu.cn

更多导师信息参见http://www.wict.pku.edu.cn/

北京大学王选计算机研究所

2021年5月26日

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **指导教师** | **职称** | **研究课题名称** | **内容简介** | **基金名称** | **联系电话** | **邮箱地址** |
| 1 | **连宙辉** | 副教授 | **汉字书法自动生成技术研究** | 本课题将应用和改进当前最先进的机器学习、人工智能、计算机视觉、计算机图形学技术开发出一个汉字书法自动生成系统。在线阶段，用户提供任一书法家的少量书法作品作为输入，该系统便可自动学习其书写风格和技巧，创作出具备相同风格且媲美该书法家书写功力的包含其他汉字的书法作品。 | 校长基金 | 82529245 | lianzhouhui@pku.edu.cn |
| 2 | **连宙辉** | 副教授 | **图像生成模型研究** | 面向图像生成任务探索研究新的有效的数学模型与方法。 | 校长基金 | 82529245 | lianzhouhui@pku.edu.cn |
| 3 | **杨德顺** | 副研究员 | **面向音乐表演版本检索的音乐旋律表征学习** | 一首音乐往往存在多种表演版本。例如，李健作曲刘兵作词的《传奇》，就有李健、王菲、孙楠和周华健等的表演版本。尽管这些版本在主旋律方面相同，但由于配器、表演者、表演时间和环境等的不同，它们在其他方面可以有很大不同。所谓音乐表演版本检索，是指给定一段音乐表演录音，自动从已知音乐表演版本库中快速找出与其主旋律相同的已知音乐表演版本。本项目的基本思路是，基于大量的有音乐ID标注的表演实例，以及少量的有音高和旋律标注的音乐表演实例，通过深度学习方法，构建基于音频的音乐旋律特征提取模型，用该模型提取的旋律特征将能准确度量音乐表演实例之间在主旋律上的相似度。 | 校长基金 | 82529553 | yangdeshun@pku.edu.cn |
| 4 | **高良才** | 副教授 | **基于深度学习的文档图像超分辨率技术** | 图像的超分辨率（SR, Super-Resolution）技术已经得到广泛的应用，文档图像的超分辨率在文档图片传输、手机拍照去模糊处理、珍贵资料复原、文字识别（OCR）等领域具有重要的研究意义和应用价值。本项目将以文本图像为研究对象，基于深度学习的理论与技术，研究更加契合文本图像的超分辨率技术，在充分考虑文档图像特征的基础上，保证文档信息超分辨率的同时降低纹理、模糊等噪声的影响，实现有针对性的超分辨率优化，相比较于传统的图像SR方法更加关注文本可读性和上下文信息，更加关注对文字识别的效果影响。在其中将研究和实现对偶学习、正则化处理、无监督去噪、背景纹理噪声的分离、感知损失的预训练在文本图像超分辨率中的应用与创新。将涉及主流的图像超分辨率技术，例如基于对偶回归网络的超分辨率重建(ESADRNet)、融合注意力的生成式对抗网络(SRGAN-SE)、分组特征度量的MRN模型、基于稀疏表示的超分网络等。 | 校长基金 | 82529655 | gaoliangcai@pku.edu.cn |
| 5 | **高良才** | 副教授 | **基于深度学习的扭曲文档自动矫正技术** | 扭曲文档矫正技术在智能判卷、拍照搜题、在线教育、文档数字化和档案信息化等领域具有着重要的应用价值；文档扭曲、折皱等形变，对文字识别（OCR）、版面理解、信息抽取等有着严重的影响，是文档文字识别领域的重要研究问题。本项目以扭曲文档图片为研究对象，研究基于深度学习的端到端文档展平拉伸模型。以扭曲的文档图片为输入，通过神经网络模型，学习扭曲文档图片与展平文档图片的映射关系，实现扭曲文档图片的矫正工作。本项目将涉及U-Net等深度神经网络、图像语义分割模型、注意力机制、Encoder-Decoder模型、数字图像处理（二值化、去噪等）、几何校正等技术，并将其应用到扭曲文档展平的工作中。 | 校长基金 | 82529655 | gaoliangcai@pku.edu.cn |
| 6 | **刘洋** | 研究员 | **基于深度学习的人与物体交互关系检测** | 理解场景不仅需要检测单个物体实例，还需要识别多个物体之间的视觉关系，其中尤为重要的一环是检测并识别每个人如何与周围物体交互，即人-物交互（HOI）检测。该任务不仅要求定位图像中的人和物体，还需要推理出人和物体之间的交互关系，比如 “吃苹果”“骑自行车”等。该任务技术的突破具有重要的科学价值，如监控视频中对人的行为分析、针对视觉障碍者的辅助技术等。然而，由于图像可能包含多个人做同一类交互，一个人同时交互多个物体、多个人共享同一个交互物体，还可能存在细粒度交互的情况，导致HOI检测颇具挑战。本项目旨在基于深度学习进行端到端的学习，以应对如上所述的一系列挑战，提高图像和视频中人与物体交互检测的精度和模型的泛化性能。 | 校长基金 | 82529250 | yangliu@pku.edu.cn |
| 7 | **刘洋** | 研究员 | **基于无监督学习的视频语义定位关键技术研究** |  视频语义定位旨在给定一个文本查询句子和一个视频，在视频中找到与对句子描述语义最匹配的视频片段（起始和结束位置）。这个任务涉及多模态信息（视觉和语言）的交互。比起视频中的动作识别和检测，由于句子可以描述更细粒度的语义信息，且不能事先定义有哪些类别，基于句子的视频语义定位难度更大。然而目前算法依赖海量的事件始末标注，限制了数据集的规模和相关算法的泛化性能。本项目旨在研究基于无监督学习的视频语义定位技术，从而大幅减少模型训练对数据标注的依赖，提高视频语义定位模型的精度和泛化能力。  | 校长基金 | 82529250 | yangliu@pku.edu.cn |
| 8 | **刘洋** | 研究员 | **视觉常识推理关键技术研究** |  视觉常识推理，指基于视觉内容对自然语言问题进行自动回答，并指出选择该答案的证据或推理依据。 视觉常识推理在一定程度上需要多模态联合学习（视觉+语言）和推理算法，实现从识别到认知的跨越，是人工智能领域的前沿热点问题。视觉常识推理模型通常可分为三个步骤，（1）视觉定位，建立自然语言问题与图像或视频间的联系，理解问题和视觉内容的含义；（2）结合图像、问题进行进一步理解，如弄清楚指代对象等,并选择问题的答案；（3）推理视觉区域之间的内在联系，并找到支持所选答案的证据和推理依据。本项目旨在基于深度学习，端到端的训练上述三个步骤，使模型拥有预测正确答案的能力, 同时提供令人信服的推理路径，以增加视觉问答系统的可解释性。 | 校长基金 | 82529250 | yangliu@pku.edu.cn |
| 9 | **王勇涛** | 副研究员 | **目标检测网络架构搜索** | 基于神经网络架构搜索的目标检测问题是当前计算机视觉、模式识别领域和深度学习领域亟需探索的新问题。本项目将系统地研究该问题，开展高精度目标检测网络架构搜索方法、面向硬件特性的目标检测网络架构搜索方法、目标检测训练数据增广策略搜索方法。本项目研究成果将同时推动目标检测和神经网络架构搜索两个问题的研究进展，具有十分重要的研究意义和应用价值。 | 校长基金 | 82529542 | wyt@pku.edu.cn |
| 10 | **赵东岩** | 研究员 | **多模态对话技术研究** | 在现实场景中，人与人之间的交流不仅限于文本形式，这个过程调动了人的各个感官来获取不同模态的信息。因此在设计人机对话系统，我们需要将不同模态的信息融合进对话生成模型来获得接近人类的对话交互。本项目将使用基于图的方法实现对多模态知识的理解与推理。通过构建知识图的方式将不同模态的知识相融合，并在此基础上使用图神经网络(GNN)等方法实现推理。此外，针对多模态对话的可解释性，我们将设计一套用于描述多模态理解和生成过程的“语言”系统，该系统基于神经模块网络(NMN)为多模态信息设计一系列基本单元，然后使用强化学习的方法对这些基本单元进行优化组合来生成灵活丰富的结果。 | 校长基金 | 82529252 | zhaody@pku.edu.cn |