

主办 CCF 计算机视觉专业委员会

COMPUTER  
VISION  
NEWSLETTER

# CCCF 计算机视觉 专委会简报

01 2021

总第 27 期



CCF 计算机视觉  
专委会

# COMPUTER VISION NEWSLETTER



## 计算机视觉专委会 简报

2021 年第 01 期

总第 27 期

### 主 办 编委会

CCF 计算机视觉专业委员会



CCF 计算机视觉  
专 委 会

#### /专委动态/

荣誉主编 **王 亮** 中国科学院自动化研究所  
主 编 **马占宇** 北京邮电大学  
执行主编 **李实英** 上海科技大学  
主 编 **毋立芳** 北京工业大学  
编 委 **黄 岩** 中国科学院自动化研究所

#### /科技前沿/

**任传贤** 中山大学  
**杨巨峰** 南开大学  
主 编 **王金甲** 燕山大学  
编 委 **储 珺** 南昌航空大学  
**崔海楠** 中国科学院自动化研究所  
**魏秀参** 南京理工大学

#### /委员风采/

主 编 **余 焯** 合肥工业大学  
编 委 **刘海波** 哈尔滨工程大学  
**赵振兵** 华北电力大学

#### /学术资源/

主 编 **李 策** 兰州理工大学  
编 委 **樊 鑫** 大连理工大学  
**贾 同** 东北大学  
**沈沛意** 西安电子科技大学

#### /海外学者/

主 编 **金 鑫** 北京电子科技学院  
编 委 **刘帅奇** 河北大学  
**张汗灵** 湖南大学

#### /视界专访/

主 编 **张军平** 复旦大学  
编 委 **贾熹滨** 北京工业大学  
**明 悦** 北京邮电大学

# CONTENTS

## 简报目录

### | 专委动态

- 04 CCF-CV 走进高校系列报告会
- 05 CCF-CV 视界无限系列研讨会
- 07 CCF-CV 荣获 2020 年度优秀专委会
- 08 2022 年度 CCF-CV 秘书处第一次工作会议召开

### | 科技前沿

- 09 无参考图像质量评价研究进展
- 12 人脸微表情检测前沿
- 14 基于深度哈希的大规模细粒度图像检索
- 16 AAAI 2021

### | 委员风采

- 19 北京科技大学殷绪成教授访谈
- 23 委员好消息

### | 学术资源

- 24 多器官分割开源代码
- 26 基于 RGB-D 的显著性物体检测数据集
- 29 好文推荐

### | 海外学者

- 32 征文通知

### | 视界专访

- 33 北京交通大学袁保宗教授专访

CCF 计算机视觉  
专委会

 CCFCV.CCF.ORG.CN

 CCFCVN@GMail.com

## CCF-CV 走进高校系列报告会

## 第 96 期 北京航空航天大学



2020年12月12日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办，北京航空航天大学承办的 CCF-CV 走进高校系列报告会第 96 期活动，通过线上直播的方式成功举行。本期活动共邀请了 5 位嘉宾做特邀报告，他们是：澳大利亚阿德莱德大学沈春华教授、北京大学林宙辰教授、西北工业大学程焱研究员、商汤科技研究执行总监代季峰博士和电子科技大学李文教授。北京航空航天大学软件学院的于茜副教授担任本次报告会的执行主席和主持人。

报告会开始由北京航空航天大学软件学院胡春明院长致欢迎辞。胡院长首先对五位报告嘉宾表示热烈的欢迎和衷心的感谢。随后，他简要介绍了北航软件学院的发展情况，并指出计算机视觉技术与软件工程的密切关系。最后，胡院长表达了对活动的祝愿以及希望未来进一步推进学院老师学生与专家学者们的深入交流。随后，本次报告会的主持人于茜老师向听众介绍了参加本次活动几位报告嘉宾的基本信息，并说明了组织本次活动的初衷，即让在校学生接触到计算机视觉和机器学习的前沿问题，拉近学生和资深专家学者的距离。



最后，她再次对来自全国各地的专家学者及中国计算机学会计算机视觉专委会表示衷心感谢。

北京大学林宙辰教授的报告题目是“Training Neural Networks by Lifted Proximal Operator Machines”。商汤科技研究执行总监代季峰博士为大家带来了题为“VL-BERT: Pre-training of Generic Visual-Linguistic Representations”的报告。西北工业大学程焱研究员的报告题目为“遥感影像‘深度’解译”，介绍了其所在团队近年来在遥感图像领域获得的研究成果。澳大利亚阿德莱德大学沈春华教授做了题为“Instance Segmentation Made Simple”的报告。电子科技大学的李文教授以“面向自动驾驶场景的跨域图像生成”为主题做了报告。

报告会最后，主持人于茜老师对报告会进行总结发言。于老师首先对进行报告的各位嘉宾以及 CCF-CV 专委会表示衷心的感谢，并希望以本次报告会为契机，欢迎全国计算机视觉领域的同行们来北京航空航天大学交流指导，拉近在校学生和专家学者之间的距离，为计算机视觉的发展持续地注入新鲜活力！

责任编辑 毋立芳

## 第 8 期 底层视觉的前沿进展与未来趋势

## CCF-CV 视界无限系列研讨会



2021年1月9日,由中国计算机学会计算机视觉专委会主办的第8期CCF-CV“视界无限”系列活动——“底层视觉的前沿进展与未来趋势”研讨会在线上成功举办,南京理工大学潘金山教授担任执行主席。研讨会邀请了计算机视觉专委会主任、北京大学查红彬教授致辞,谷歌资深科学家 Deqing Sun 博士、北京理工大学付莹教授、新加坡南洋理工大学 Chen Change Loy 副教授、香港科技大学陈启峰助理教授以及哈尔滨工业大学左旺孟教授做主题报告。美国加州大学默塞德分校 Ming-Hsuan Yang 教授及以上五位讲者参与了圆桌讨论。计算机视觉专委会 B 站公众号对本次会议进行了全程直播,直播人气峰值达到 5400+。

首先,查红彬教授代表计算机视觉专委会致辞。查老师指出“视界无限”活动聚焦于比较细节的视觉问题,参与者对这些问题进行深入研讨,大家可以借此了解最新的研究进展,探索和发现一些潜在的兴趣点。查老师指出底层视觉是计算机视觉整个研究链中相对靠近底端的部分,其中有很多问题是相当重要的,同时也极具挑战性。他分析了本次活动的两个重要意义:其一,底层视觉问题和高层视觉问题是密切相关的,很多高层视

觉返回的信息能够帮助解决底层视觉问题,从这一方面说,它不单纯是视觉图像处理这么纯粹的事情,还和高层处理有密切联系;其二,人的视觉系统有非常强的底层处理能力,脑的处理皮层里面具有相当高效的底层处理功能,这些功能是局部化的,并具有并行处理的机制,其中一些机制也为我们在计算机视觉中处理底层问题提供了很好的研究线索,因此借助大脑的相关机理能够为解决底层视觉问题提供一些新的思路。最后,查老师简要介绍了本次研讨会的讲者情况,并预祝本次视界无限活动能够圆满成功!



Deqing Sun 博士报告的主题是“Learning Optical Flow: From MRFs to CNNs”。他指出,世界是动态的,想要理解动态的世界,必须要理解运动。光流是描述运动的有效方式,其作为一个重要问题,面临着许多挑战,如大幅度位移、运动模糊、无纹理区域、遮挡、光源变化、噪声等。因此为光流的形成过程和先验知识选取合适的描述方法是十分重要的。Deqing Sun 博士在报告中介绍了光流问题两种不同方向的工作。首先是基于 MRF (马尔可夫随机场) 和 CRF (非局部条件随机场) 的方法;其次是结合光流领域知识和 CNNs (卷积神经网络) 的方法。前者可以得到较好的

结果，但是计算量非常大，且很难处理复杂场景，如遮挡，光源变化；后者可以达到实时计算，但往往较难达到非常好的性能。为了同时获得基于 CNNs 方法的实时计算和传统方法的优越性能，Deqing Sun 博士提出了 PWC-Net 方法，其利用多尺度的 cost-volume 来解决光晕问题，并借助频繁使用光流领域知识来避免引入庞大的计算量，在使用更少的参数量的情况下达到了更好效果。最后，Deqing Sun 博士介绍了光流领域最近的进展，并分享了研究光流问题时的一些经验和建议。

付莹教授的报告题目是“噪声建模与图像重构”，如何在极端低光的情况对条状等噪声去噪进行了细致的介绍。以往方法大多需要图像序列或者成对的训练数据集，获得难度较大。针对上述问题，付莹教授提出直接对噪声进行建模来合成与真实场景相匹配的噪声图像，并提出了一个噪声参数标定的方法，以便模型适用于各类给定的相机。针对图像复原的问题，付莹教授采用了 plug-and-play 的方法。对于其对参数敏感的问题，给出了解决方案。

吕健勤 (Chen Change Loy) 教授的报告题目是“Deep Generative Prior”。吕教授从图像复原问题出发，指出解决图像复原问题需要有一个好的图像先验。他举例说明了已有的深度先验方法的一些缺点，介绍了两方面的工作：(1) 如何利用已经训练好的 GAN 作为图像先验。(2) 将训好的 GAN 拆入到 Encoder-Decoder 结构中来恢复图像。他指出，GAN 训练 generator 可以作为一个比较通用的先验来实现各种图像复原任务。

陈启峰教授的报告题目是“Learning-based Sensing Technologies”。陈教授分析了现有 RGB 摄像头、深度摄像头以及 LiDAR 传感器存在的问题。针对这些问题，他介绍了在极暗条件下如何从丰富的 raw data 信息中复原出更清晰的图像。其次，介绍了如何通过长焦短焦镜头获取 raw data 训练数据用于解决图像超分辨率问题。然后，陈启峰教授设计 two-stage 的方案解决了 reflection removal 的问题，第一个 stage 引入 polarization 得到较好的 reflection 的信息，基于此，第二个 stage 恢复出 transmission 的信息。最后，陈

教授介绍了如何通过多个摄像头的设计拍到更远处 (300 米-500 米) 的深度图以及边缘更丰富的深度图。

左旺孟教授的报告题目是“自监督上下文建模及底层视觉应用”。左教授在报告中深入浅出、细致全面地介绍了自监督学习在底层计算机视觉的深度上下文建模，图像压缩和图像去噪这三方面取得的研究进展。

在 Panel 环节，与会嘉宾就“底层视觉中有哪些值得关注的问题？其主要挑战性和发展趋势是什么”、

“目前深度学习方法被广泛应用于底层视觉中的大多数问题，比如图像超分辨率、图像去噪、图像去模糊等，但是目前大多数方法过分依赖数据，如何摆脱对训练数据的过分拟合而具有更强的泛化和外推能力”、“目前基于深度学习方法取得了显著的进展，在这类方法中还有哪些可以值得研究的方向，比如网络设计需考虑哪些方面？传统的先验建模方法是否还值得研究”、“不同于其他高层视觉问题，在图像复原等相关任务中，我们没法从退化图像中人工标注出清晰数据，目前大多通过特定的数据获取方法来获得训练数据。在算法层面上也出现了大量的 Self-supervised learning 或者 Unpaired learning 的方法，如何看待这些新方法”、

“目前基于 Transformer 的方法受到大量的关注，如何看待 Transformer 方法在底层视觉问题上的应用”、

“高层视觉任务中的大多数问题需要考虑相关的语义信息。近年来也出现了大量借助于高层视觉任务中的方法解决底层视觉任务的尝试，并且取得了较好的效果。那么在底层视觉任务中的一些问题，比如图像复原，是否有助于高层视觉任务？”等问题展开热烈讨论。



最后，第 8 期“视界无限”研讨会在中午 12 点 10 分圆满结束。

责任编辑 杨巨峰

## CCF-CV 荣获 2020 年度优秀专委会



中国计算机学会“CCF 专委年度总结交流评估会”于 1 月 9 日-10 日在北京友谊宾馆召开。CCF-CV 专委主任查红彬教授参加会议并汇报工作。CCF 所属 38 个专委会参与了年度工作评估，评估组依据专委汇报内容以及专委和学会秘书处提供的相关材料，对 2020 年度各专委的“工作及学术活动的规范性”、“对学会工作的参与及贡献”和“学术水平及影响力”等方面进行了评估。综合专委互评和联系会员评价的成绩，CCF-CV 专委荣获 2020 年度优秀专委会！

CCF-CV 专委会成立于 2013 年，是直属于中国计算机学会的计算机视觉领域的专业分支机构。专委会的目标是就计算机视觉学科的专业内容更好地开展学术/技术交流、发展战略研究，促进国内学者间的了解与合

作，推动国内计算机视觉学科发展，提升我国计算机视觉研究在国际领域的影响力。成立至今曾多次获得学会颁发的“特色活动奖”、“综合进步奖”、“优秀专委会”、“年度特别奖”。

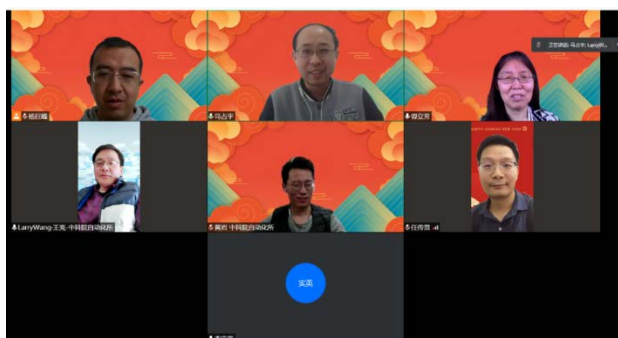


感谢大家过去一年对 CCF-CV 专委工作的大力支持和帮助！新的一年让我们继续努力，再创辉煌！

责任编辑 黄岩

## 2021 年度 CCF-CV 秘书处第一次工作会议召开

2021 年 2 月 20 日，中国计算机学会计算机视觉专委会（CCF-CV）秘书处本年度第一次工作会议于线上召开。专委会副主任王亮研究员参会指导工作，秘书处全体成员参加了会议，会议由秘书长马占宇教授主持。



本次会议主要讨论和制定秘书处 2021 年度工作计划。过去一年，专委会各项学术活动和组织建设稳步推进，走进高校、走进企业、视界无限、专委简报与

网站的活动形式与内容不断推陈出新，微信公众号粉丝突破一万，继续展示 CCF-CV 专委会的组织活力与特色，专委会也获得计算机学会优秀专委会称号。新的一年，各项活动如何更有实效地开展，如何科学有效地统计委员贡献，如何更广泛地激发委员们参与专委会建设的积极性，如何办好年度学术论坛 RACV 2021，大家围绕这些议题展开了热烈讨论，并形成了具体可行的执行方案。

专委会品牌活动 RACV2021 已进入筹备阶段，秘书处将积极配合组委会做好组织和服务工作；继续丰富与企业的交流形式，拓展交流渠道；进一步调动委员的参与度，广泛采纳委员建议，对有突出贡献的委员和志愿服务人员进行表彰奖励。

责任编辑 黄岩

热点追踪

## 无参考图像质量评价研究进展

江西财经大学 方玉明

## 一、引言

近年来，图像数据呈爆发式增长，图像的视觉质量与用户体验密切相关，是决定计算机视觉应用的重要因素之一。不同的内在因素如成像设备老化等<sup>[1]</sup>和外在因素如拍摄环境等都与图像质量紧密相关。

图像质量评价(Image Quality Assessment, IQA)是指通过对图像信号进行相关特性分析，量化图像的视觉失真程度。评价方法主要包括主观质量评价和客观质量评价。主观质量评价指的是由主观判断获取图像视觉质量；客观质量评价指的是通过客观算法自动计算图像的视觉质量。根据计算图像视觉质量时是否需要参考图像的信息，客观质量评价可分成三类：全参考质量评价(Full Reference IQA, FR-IQA)；半参考质量评价(Reduced Reference IQA, RR-IQA)和无参考质量评价(No Reference IQA, NR-IQA)。FR-IQA模型和RR-IQA模型通过对图像的视觉特征进行分析，量化参考图像和失真图像之间的差异，计算失真图像的视觉质量。相对于FR-IQA模型和RR-IQA模型，NR-IQA模型在计算失真图像的视觉质量时不需要任何参考图像的信息，在实际应用系统中具有更广泛的应用前景。

本文重点介绍 NR-IQA 模型的相关研究进展和未来发展趋势，内容框架如图 1 所示。根据 NR-IQA 模型训练时是否需要图像的主观分数，将 NR-IQA 模型分为有监督学习的无参考模型(需要主观分数)和无监督学习的无参考模型(不需要主观分数)，并对每类模型从基于传统机器学习的模型和基于深度学习的模型两个方面进行梳理。

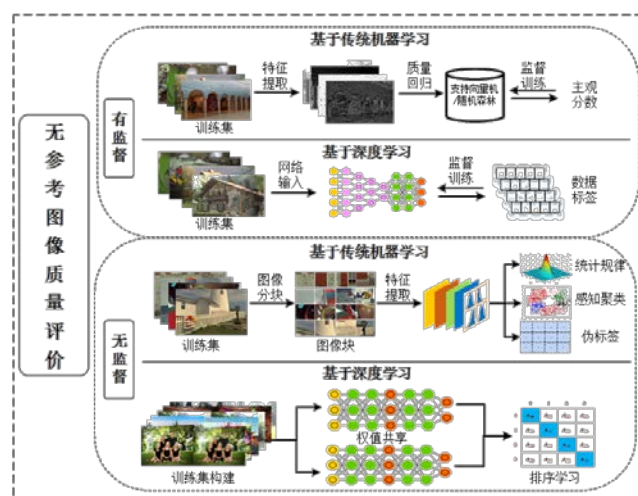


图 1 本文内容框架

## 二、研究现状

## 2.1. 有监督学习的 NR-IQA 模型

有监督学习的 NR-IQA 模型指的是模型训练过程中需要用到主观数据，模型的输入为图像或者图像特征，输出为主观分数，通过迭代训练，获得最终的 NR-IQA 模型。有监督学习的 NR-IQA 模型包括基于传统机器学习算法的有监督学习模型和基于深度学习算法的有监督学习模型。前者依赖先验知识手动提取视觉特征，后者通过端到端的方式学习特征表达。

基于传统机器学习的有监督学习 NR-IQA 模型旨在设计有效的视觉特征表达方法，能较好地反映图像视觉质量变化，并通过传统机器学习算法如支持向量机(Support Vector Machine, SVM)或随机森林(Random Forest, RF)等学习视觉特征到图像视觉质量的映射模型，此类模型的研究重点在于特征提取。Mittal 等人<sup>[2]</sup>较早提出使用空域下的自然场景统计(Natural

Scene Statistics, NSS)特征来构建 IQA 模型。与此方法<sup>[2]</sup>的设计思路类似, Fang 等人<sup>[3]</sup>研究发现图像的矩特征和信息熵呈现某种分布,并提出使用拟合得到的分布函数的输出值作为失真图像的特征,以构建 NR-IQA 模型。相较于 Mittal 等人提出的方法<sup>[2]</sup>特征提取过程需要拟合函数分布, Wu 等人<sup>[4]</sup>则设计了特征提取过程不需要拟合函数分布的 NR-IQA 模型。

不同于传统手工提取特征的 NR-IQA 模型,基于深度学习的模型可以端到端学习图像与图像质量之间的映射关系,且性能远超过基于传统机器学习的模型,该类模型的重点在于卷积神经网络(Convolutional Neural Network, CNN)的设计。除了直接端到端学习图像至质量分数之间映射的模型<sup>[5,6]</sup>外,研究人员也设计了包含多个模块的模型<sup>[7]</sup>。为提高 NR-IQA 模型性能, Ma 等人<sup>[8]</sup>将 NR-IQA 任务分成多个子任务,利用其他任务辅助 IQA 任务。总体而言,基于深度学习的 NR-IQA 模型在大多数数据库上的表现都超过了基于传统机器学习的模型,性能接近饱和。

## 2.2. 无监督学习的 NR-IQA 模型

无监督学习的 NR-IQA 模型指的是模型训练过程中不需要主观数据。类似有监督学习的 NR-IQA 模型,无监督学习的 NR-IQA 模型,主要包括基于传统机器学习的模型和基于深度学习的模型。

基于传统机器学习算法的无监督 NR-IQA 模型通过构建统计模型度量图像质量的差异,基于深度学习的无监督 NR-IQA 模型通过构造有效训练数据,构建神经网络模拟排序信息的学习过程,实现图像与质量分数之间的映射。该类方法一定程度上解决了数据驱动模型中训练数据匮乏的问题,且性能远高于基于传统机器学习的 NR-IQA 模型。如 Ma 等人<sup>[9]</sup>提出一个基于多标签学习的 NR-IQA 模型,通过对收集的高清图像添加若干种类的人工合成失真得到失真图像数据集,同时利用若干种现有 IQA 方法对失真图像数据集进行质量分数预测,基于图像对生成机制生成图像对,通过两两比较若干 IQA 方法计算的质量分数得到图像对对应标签。该方法提出的训练图像数据扩增方法,在一定程度上解决了现有 IQA 模型因图像数据不足而可能导致模型过拟合的

问题,且所提算法可高效结合现有 IQA 方法,模型简单有效,可扩展能力强。

与基于传统机器学习的有监督 NR-IQA 模型类似,基于传统机器学习的无监督 NR-IQA 模型同样依赖手工特征提取,其性能也受到特征表达能力的影响。同时,常用的使用 FR-IQA 模型获取图像伪标签的方式存在明显缺陷。另外,有限的训练数据内容也是影响这类模型可扩展能力的关键因素。基于深度学习的无监督 NR-IQA 模型可以充分利用海量的无标签数据,通过对比的方式或者借用 FR-IQA 模型产生伪标签,用于模型训练,可有效提升算法性能和算法扩展性,值得进一步研究。

## 2.3. 数据库

自然图像数据库可分为合成失真数据库和真实失真数据库两部分。常用的合成失真数据库包括 LIVE(Laboratory for Image and Video Engineering)数据库<sup>[10]</sup>和 TID2013(Tampere Image Database 2013)数据库<sup>[11]</sup>等。随着对真实失真应用场景研究的不断深入,一系列真实失真数据库不断涌现,包括 LIVE Challenge 数据库<sup>[12]</sup>和 KonIQ-10k(Konstanz Authentic Image Quality Database)数据库<sup>[13]</sup>等。

## 三、未来发展方向

无参考图像质量评价在评价内容和客观模型方面,均存在潜在的发展方向。就评价内容而言,现有的 IQA 数据库大多通过人工添加常见噪声生成,且获取的失真图像内容形式较为单一,这些失真图像仅仅来源于数量有限的高质量图像。在真实应用系统中,需要处理的图像/视频数据内容形式复杂多样。并且,这些数据在生成、传输、处理等过程中可能经历了多重失真,引入的失真无法完全使用计算机进行模拟生成。这不仅限制了现有合成失真数据库的应用价值,还限制了针对人工添加噪声数据设计的客观评价模型的可扩展性。未来仍需要投入大量精力对真实失真图像、特定应用场景下的失真图像数据进行研究。虽然陆续提出了自然失真图像质量评价的解决方案,该类研究仍然具有较大的挑战性,提升空间依然较大。

责任编辑 储珺

## 参考文献

- [1] Fang Y M, Zhu H W, Zeng Y, Ma K D and Wang Z. 2020. Perceptual quality assessment of smartphone photography//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 3677-3686 [DOI: 10.1109/CVPR42600.2020.00373]
- [2] Mittal A, Moorthy A K and Bovik A C. 2012. No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing, 21(12): 4695-4708 [DOI: 10.1109/TIP.2012.2214050]
- [3] Fang Y M, Ma K D, Wang Z, Lin W S, Fang Z J and Zhai G T. 2015. No-reference quality assessment of contrast-distorted images based on natural scene statistics. IEEE Signal Processing Letters, 22(7): 838-842 [DOI: 10.1109/LSP.2014.2372333]
- [4] Wu J J, Zhang M, Li L D, Dong W S, Shi G M and Lin W S. 2019. No-reference image quality assessment with visual pattern degradation. Information Sciences, 504: 487-500 [DOI: 10.1016/j.ins.2019.07.061]
- [5] Kang L, Ye P, Li Y and Doermann D. 2014. Convolutional neural networks for no-reference image quality assessment//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE: 1733-1740 [DOI: 10.1109/CVPR.2014.224]
- [6] Kim J, Nguyen A D and Lee S. 2019. Deep CNN-based blind image quality predictor. IEEE Transactions on Neural Networks and Learning Systems, 30(1): 11-24 [DOI: 10.1109/TNNLS.2018.2829819]
- [7] Pan D, Shi P, Hou M, Ying Z F, Fu S Z and Zhang Y. 2018. Blind predicting similar quality map for image quality assessment//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE: 6373-6382 [DOI: 10.1109/CVPR.2018.00667]
- [8] Ma K D, Liu W T, Zhang K, Duanmu Z F, Wang Z and Zuo W M. 2018. End-to-end blind image quality assessment using deep neural networks. IEEE Transactions on Image Processing, 27(3): 1202-1213 [DOI: 10.1109/TIP.2017.2774045]
- [9] Ma K D, Liu X L, Fang Y M and Simoncelli E P. 2019. Blind image quality assessment by learning from multiple annotators//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 2344-2348 [DOI: 10.1109/ICIP.2019.8803390]
- [10] Sheikh H R, Wang Z, Cormack L and Bovik A C. 2006. LIVE image quality assessment database release 2[DB/OL]. [2020-07-08]. <https://live.ece.utexas.edu/research/quality/>
- [11] Ponomarenko N, Jin L N, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F and Jay Kuo C C. 2015. Image database TID2013: peculiarities, results and perspectives. Signal Processing: Image Communication, 30: 57-77 [DOI: 10.1016/j.image.2014.10.009]
- [12] Ghadiyaram D and Bovik A C. 2016. Massive online crowdsourced study of subjective and objective picture quality. IEEE Transactions on Image Processing, 25(1): 372-387 [DOI: 10.1109/TIP.2015.2500021]
- [13] Hosu V, Lin H H, Sziranyi T and Saupe D. 2020. KonIQ-10k: an ecologically valid database for deep learning of blind image quality assessment. IEEE Transactions on Image Processing, 29: 4041-4056 [DOI: 10.1109/TIP.2020.2967829]



## 方玉明

江西财经大学教授。主要研究方向为计算机视觉、多媒体信号处理、视觉质量评估。担任中国计算机学会计算机视觉专委会委员。

E-mail: leo.fangyuming@foxmail.com

热点追踪

## 人脸微表情检测前沿

中国科学院心理研究所 李婧婷 王甦菁

**微**表情是一种短暂的、无意识的面部表情，通常出现在个体试图隐藏真实感受的时刻。微表情的分析有很多潜在的应用价值，例如在医疗关怀、执法审讯、国家安全等领域。和常见的普通表情相比，微表情有三个显著特征：持续时间短(<500ms)、强度低和局部运动。人类用肉眼很难发现和识别如此短暂而微弱的表情。因此需要借助计算机视觉来进行相关的分析。其中，微表情检测的目标是定位微表情在视频中发生的时刻，这一研究非常具有实用价值。如果能在一段视频中准确地检测和定位到某个时间点有微表情出现，那么就说明这个人在这个时刻可能会有异常。然而，在真实场景下，从大量头部动作和人脸表情中检测分离出微小短暂的微表情是极具挑战性的任务。因此，目前微表情检测的研究相对较少。同时，由于微表情数据库的样本量较少，限制了深度学习技术与微表情检测方法的结合。近年来，中国科学院心理研究所发布了更适合用于微表情检测研究的长视频数据集 CAS(ME)<sup>2</sup>。研究长视频中微表情的检测方法将有助于实现微表情分析在真实场景的应用。

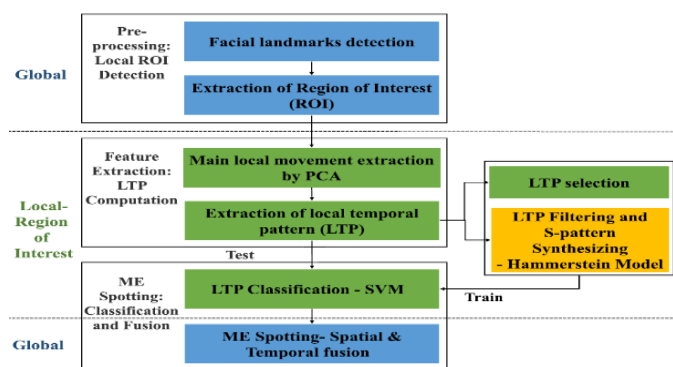


图1 结合局部时空特征提取、局部机器学习和晚期融合的微表情检测流程

针对微表情持续时间短、动作强度低以及局部发生三个特点，提出了结合局部时空特征(LTP)提取、局部机器学习和晚期融合的微表情检测方法，如图1所示。我们提出的LTP特征针对微表情的运动具有特定的变化曲线(S-模式)，如图2所示。通过使用机器学习对LTP进行分类，可以将微表情与其他面部动作区分开。此外，我们还提出了覆盖整个面部的全局融合分析，以改善检测系统对微表情(局部)运动与头部(全局)运动之间的区别能力。但是，基于S-模式的机器学习模型训练受限于微表情样本数目。由于Hammerstein模型(HM)可以有效地模拟肌肉运动模式，我们利用HM生成S-模式，扩充了用于S-模式训练的数据集，从而提升了该方法的微表情检测性能。

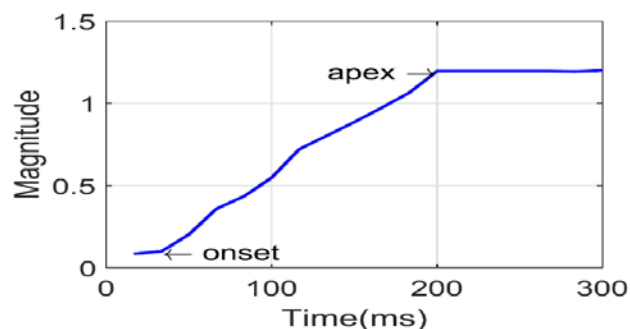


图2 针对微表情的局部时空模式(LTP)

此外，受启发于空间域上的目标检测技术，我们尝试将相应技术扩展到时间域上进行微表情检测，并提出了一个多尺度的卷积神经网络(CNN): MESNet。据我们所知，这是CNN首次与长视频中微表情检测方法相结合。基于轻量级的2+1D-时空卷积网络，如图3所示，我们增加了一个片段候选网络(Clip Proposal Network, CPN)和分类回归网络(CRNet)。其中，CPN

被添加到 2+1D-时空卷积网络预训练的卷积层中,并将所有可能的微表情区间列入候选结果,如图 4 所示。然后,CRNet 模块对候选区间进行进一步判断,并且回归它们的时间边界,从而得出最终的微表情片段检测结果,如图 5 所示。实验结果表明,我们所提出的 MESNet 算法在长视频样本中取得了最优的微表情片段检测性能。

以上工作已分别被国际期刊 IEEE Transactions on Affective Computing(TAC)和 IEEE Transactions on Image Processing(TIP)接收。

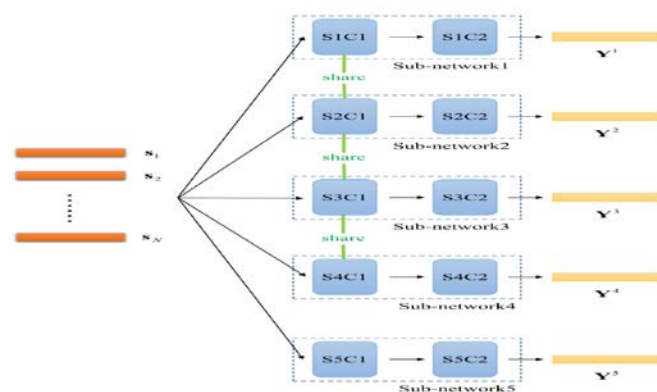


图 4 片段候选网络 (Clip Proposal Network, CPN)

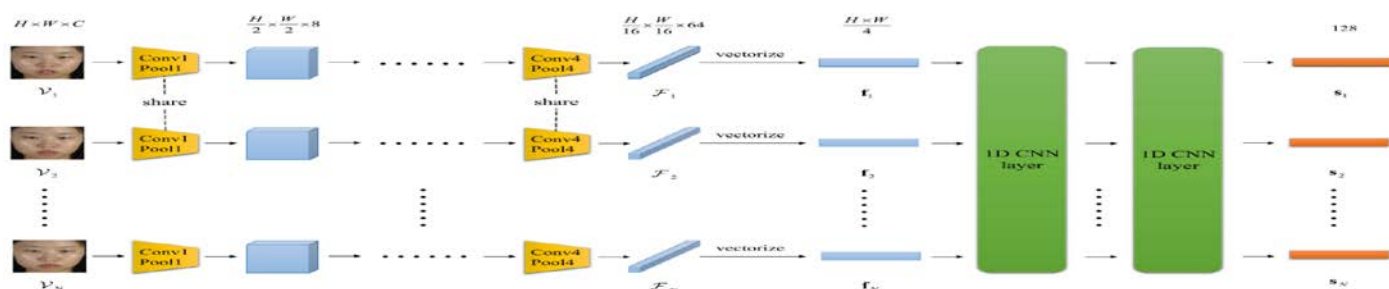


图 3 2+1D-时空卷积网络

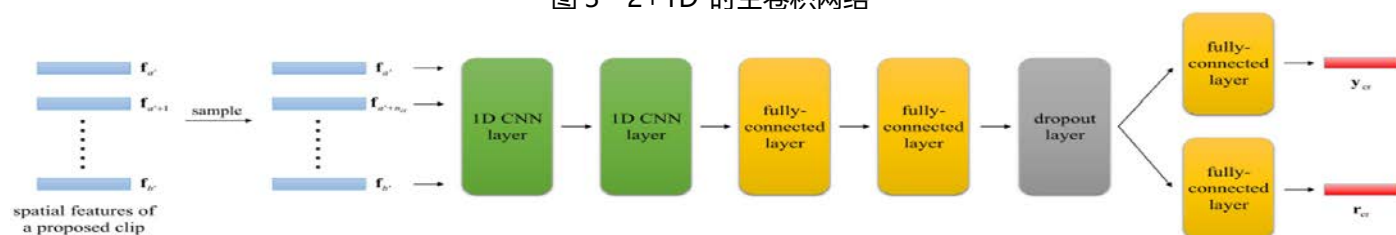


图 5 分类回归网络 (CRNet)

责任编辑 魏秀参

参考文献

- [1] Li, Jingting, Catherine Soladie, and Renaud Seguier. "Local Temporal Pattern and Data Augmentation for Micro-Expression Spotting." IEEE Transactions on Affective Computing (2020).
- [2] Wang, Su-Jing, He, Ying, Li, Jingting and Fu, Xiaolan. "MESNet : A Convolutional Neural Network for Spotting Multi-Scale Micro-Expression Intervals in Long Videos." IEEE Transactions on Image Processing (2021).



王甦菁

中国科学院心理研究所副研究员, 博士生导师。主要研究方向为模式识别与机器学习, 特别是微表情识别。在国内外重要期刊和学术会议上发表五十余篇论文, 包括 TIP、TNN、ECCV 等。获 2018 年第八届吴文俊人工智能科学技术奖一等奖。担任中国计算机学会计算机视觉专委会委员。  
Email: wangsuqing@psych.ac.cn



李婧婷

中国科学院心理研究所博士后。主要研究方向为图像处理与计算机视觉, 特别是微表情检测与识别。Email: lijting@psych.ac.cn

热点追踪

# 基于深度哈希的大规模细粒度图像检索

南京理工大学 魏秀参 沈阳

随着网络图像数据的快速增长，在面向大规模图像检索时，图像表示学习与哈希学习相结合的深度哈希技术已成为发展趋势。从大规模数据集中检索图像会面临查询速度和高冗余存储成本两个问题，而细粒度图像的细微视觉差距更是加剧了图片识别的难度。为缓解这些技术难题，我们研究了基于深度哈希的细粒度图像检索主题，通过局部和全局特征的提取并在此基础上进行局部特征对齐与交换，保证了局部特征在图像间的识别能力和语义一致性，为细粒度图像生成紧凑的二进制哈希编码。此外，与其他最近邻方法相比，提出的方法在速度与存储空间上也有较大的改善。相关成果被 ECCV 2020 收录为大会口头报告论文。

得益于深度学习发展，哈希学习已被证明是一种有成效的大规模图像检索解决方案，因它可以大大降低存储成本，提高查询速度。哈希学习作为近似最近邻(ANN)搜索代表性研究领域，近年来哈希学习已经成功应用于大规模图像检索任务。我们希望探讨哈希算法在细粒度图像检索中的有效性，这也是研究细粒度哈希问题的首项工作，该任务主要涉及如何为细粒度对象设计有效且高效的哈希编码问题。如图 1 所示，细粒度图像具有类间差异小、类内差异大的特性，细粒度哈希方法需要为具有相同子类的图像生成紧凑二进制哈希编码并为不同子类的图片生成差异较大编码。为了解决这一具有挑战性任务，我们提出了 ExchNet 网络，其框架主要由三个模块组成：(1)表示学习模块，用以获取细粒度目标的局部和全局特征；(2)局部特征对齐模块，对从不同细粒度图像中获得的局部特征进行匹配；(3)哈希码学习模块，用于生成二进制哈希编码。整体框架结构如图 2 所示。

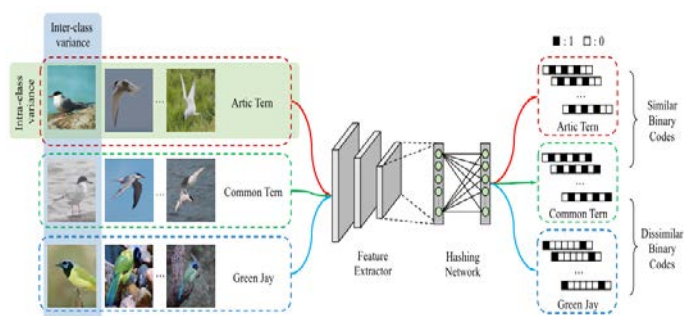


图 1 细粒度哈希任务示意图

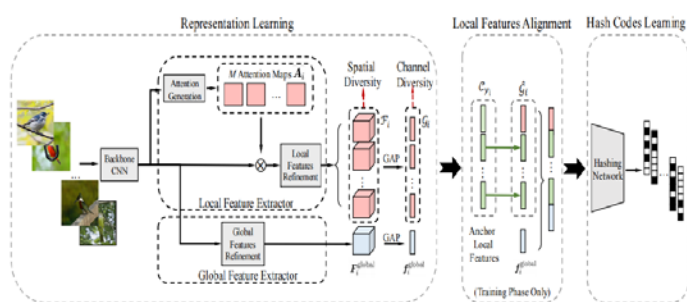


图 2 ExchNet 整体框架

在表示学习模块中，除了获取全局特征外，我们还使用注意力机制捕捉局部特征以表示图像的细粒度特征。特征对齐的方法如图 3 所示，其主要基于假设：对于两张具有相同子类的图片，交换对应细粒度对象部件(part)的局部特征(如图 3 中两只鸟的翅膀)，不会影响这两张图片哈希编码的生成以及它们的相似性。在训练过程中我们会交换同类样本的局部特征并同时保证同类样本哈希编码相似性，进而达到隐式的特征对齐目的。

为了证明提出方法的实用性和有效性，我们首先与其他 ANN 方法进行比较。在数据集 FOOD101 上，与线性搜索相比，我们的方法可以在 512 维和 1024 维的特征上分别达到 233 倍和 395 倍的加速比，同时方法

的内存开销也比基于树的方法少。在表 2 中，我们对比了 8 种经典的针对通用图像的哈希学习方法，并比较了平均精度(mean Average Precision, MAP)的检索结果，可以看出我们的方法在选取的所有细粒度数据集中均能取得最好的检索精度。

表 1 ExchNet 与 ANN 方法在 Food101 上的性能比较

Method	512-dim				1024-dim			
	P@10(↑)	WTime(↓)	Speedup(↑)	Memory(↓)	P@10(↑)	WTime(↓)	Speedup(↑)	Memory(↓)
Linear	80.05%	9,481.03	1x	207.2MB	80.28%	22,377.96	1x	414.1MB
BallTree	77.22%	236.23	40.13x	28.1MB	77.74%	213.88	104.62x	28.1MB
KDTree	77.42%	70.16	135.13x	28.8MB	77.73%	73.57	304.14x	28.7MB
PQ	77.12%	43.49	217.99x	524.5KB	77.18%	72.47	308.74x	1.0MB
Ours	77.69%	40.54	233.85x	404.0KB	78.06%	56.57	395.53x	404.0KB

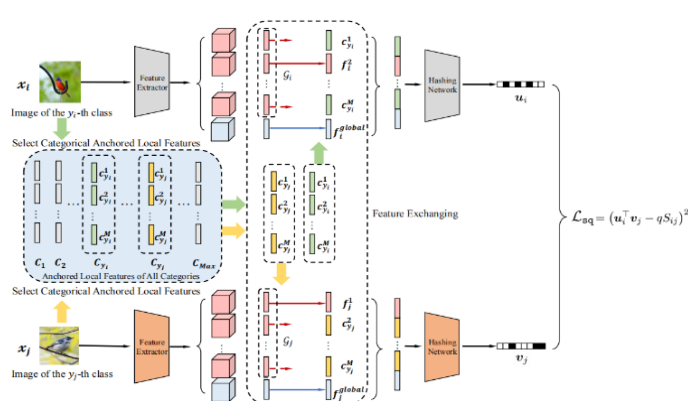


图 3 局部特征对齐方法

表 2 多个细粒度图像标准数据集的检索精度(MAP)比较

Method	#Bits	LSH	SH	ITQ	SDH	DPSH	DSH	HashNet	ADSH	Ours
CUB	12bits	2.26%	5.55%	6.80%	10.52%	8.68%	4.48%	12.03%	20.03%	<b>25.14%</b>
	24bits	3.59%	6.72%	9.42%	16.95%	12.51%	7.97%	17.77%	50.33%	<b>58.98%</b>
	32bits	5.01%	7.63%	11.19%	20.43%	12.74%	7.72%	19.93%	61.68%	<b>67.74%</b>
	48bits	6.16%	8.32%	12.45%	22.23%	15.58%	11.81%	22.13%	65.43%	<b>71.05%</b>
Aircraft	12bits	1.69%	3.28%	4.38%	4.89%	8.74%	8.14%	14.91%	15.54%	<b>33.27%</b>
	24bits	2.19%	3.85%	5.28%	6.36%	10.87%	10.66%	17.75%	23.09%	<b>45.83%</b>
	32bits	2.38%	4.04%	5.82%	6.90%	13.54%	12.21%	19.42%	30.37%	<b>51.83%</b>
	48bits	2.82%	4.28%	6.05%	7.65%	13.94%	14.45%	20.32%	50.65%	<b>59.05%</b>
NABirds	12bits	0.90%	2.12%	2.53%	3.10%	2.17%	1.56%	2.34%	2.53%	<b>5.22%</b>
	24bits	1.68%	3.14%	4.22%	6.72%	4.08%	2.33%	3.29%	8.23%	<b>15.69%</b>
	32bits	2.43%	3.71%	5.38%	8.86%	3.61%	2.44%	4.52%	14.71%	<b>21.94%</b>
	48bits	3.09%	4.05%	6.10%	10.38%	3.20%	3.42%	4.97%	25.34%	<b>34.81%</b>
VegFru	12bits	1.28%	2.36%	3.05%	5.92%	6.33%	4.60%	3.70%	8.24%	<b>23.55%</b>
	24bits	2.21%	4.04%	5.51%	11.55%	9.05%	8.91%	6.24%	24.90%	<b>35.93%</b>
	32bits	3.39%	5.65%	7.48%	14.55%	10.28%	11.23%	7.83%	36.53%	<b>48.27%</b>
	48bits	4.51%	6.56%	8.74%	16.45%	9.11%	17.12%	10.29%	55.15%	<b>69.30%</b>
Food101	12bits	1.57%	4.51%	6.46%	10.21%	11.82%	6.51%	24.42%	35.64%	<b>45.63%</b>
	24bits	2.48%	5.79%	8.20%	11.44%	13.05%	8.97%	34.48%	40.93%	<b>55.48%</b>
	32bits	2.64%	5.91%	9.70%	13.36%	16.41%	13.10%	35.90%	42.89%	<b>56.39%</b>
	48bits	3.07%	6.63%	10.07%	15.55%	20.06%	17.18%	39.65%	48.81%	<b>64.19%</b>

责任编辑 王金甲

参考文献

- [1] Q. Cui, Q.-Y. Jiang, X.-S. Wei\*, W.-J. Li, and O. Yoshie. ExchNet: A Unified Hashing Network for Large-Scale Fine-Grained Image Retrieval. European Conference on Computer Vision, Virtual Conference, 2020, pp. 189-205. (Oral Presentation).
- [2] X.-S. Wei, J.-H. Luo, J. Wu, and Z.-H. Zhou. Selective Convolutional Descriptor Aggregation for Fine-Grained Image Retrieval. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881. (ESI 高被引论文, Google Scholar Citations: 217).



魏秀参

南京理工大学计算机科学与工程学院教授。主要研究领域为计算机视觉和机器学习。担任中国计算机学会计算机视觉专委会委员。

Email: weixs@njust.edu.cn



沈阳

南京理工大学计算机科学与工程学院硕士生。主要研究领域为计算机视觉和机器学习。

Email: shenyang\_98@njust.edu.cn

顶会观察

## AAAI 2021

燕山大学 王金甲

人工智能发展协会 (Association for the Advancement of Artificial Intelligence, AAAI) (前身为美国人工智能协会 (American Association for Artificial Intelligence, AAAI)) 是人工智能(AI)领域的主要学术组织之一, 致力于促进对思维和智能行为背后机制及其在机器中体现的科学理解。第三十五届 AAAI 人工智能大会(AAAI-21)于 2021 年 2 月 2-9 日召开完全线上虚拟会议。总主席由香港科技大学杨强教授担任(这是华人首次担任此职位), 项目主席由加拿大不列颠哥伦比亚大学 Kevin Leyton-Brown 和印度德里理工学院 Mausam 担任。在中国计算机学会的国际学术会议排名中, AAAI 会议被视为人工智能领域的 A 类顶级会议, 在 2020 年谷歌学术指标总榜单排名第 96 位。

## 一、AAAI 2021 的亮点

AAAI 会议的目的是促进人工智能的研究以及 AI 研究人员, 从业人员, 科学家和所属学科的工程师之间的科学交流。与 AAAI-20 半线上会议相比, AAAI-21 会议采用了完全线上虚拟会议。每篇被接受论文的作者提前录制视频演讲并在整个会议期间提供观看, 每个与会者可以远程参与和电子互动。

AAAI-21 会议主要分为两个技术 track。主 track 包括会议大部分论文, 此外会议委员会特别鼓励并突出展示三个领域的时事话题: 神经符号 AI; AI 对新冠肺炎大流行的反应; 用于会议组织和交付的 AI。所有主 track 论文按照相同的标准和程序进行评审。次 track 聚焦人工智能对社会的影响, 其论文与主 track 论文有着不同的评审标准。

AAAI-21 会议采用两轮双盲审稿政策。第一轮审稿, 每篇论文分配两名审稿人。如果这两名审稿人认为该论文无法被录用, 那么该论文将直接被拒绝。没有被拒的论文将进入第二轮。第二轮审稿, 每篇论文分配另外两个审稿人。新的审稿人给出意见之前, 他们是无法看到第一轮的评审意见, 也就是说这两轮审稿相互独立。另外, 会议还提供了 NeurIPS 和 EMNLP 会议被拒论文的快速通道: 平均分数至少为 4.9(NeurIPS-20) 或 2.8(EMNLP-20 长论文)的被拒论文可以和先前的评审意见和论文编号等信息一起直接提交到 AAAI 的第二轮供审稿人评审。

## 二、论文录用情况

AAAI-21 会议共收到了 9034 篇论文投稿, 超过了去年的 8800 篇, 其中有效审稿为 7911 篇, 最终录取数量为 1692 篇, 接收率为 21.4%, 略高于去年的 20.6%。

据统计, AAAI-21 会议中国大陆投稿论文数量为 3319 篇, 被接收论文数量为 627 篇, 论文投稿量与接收量均为全球第一。12 篇获奖论文中有 5 篇出自华人作者团队, 包括两篇最佳论文奖论文、两篇提名奖论文以及一篇杰出论文奖论文。国内企业, 如阿里、百度、京东和腾讯分别有 54、24、21 和 11 篇论文被收录。国内高校与企业人工智能领域的影响力日趋扩大。

AAAI-21 会议接受内容包括机器学习(深度学习、统计学习等)、自然语言处理、计算机视觉、数据挖掘、多智能体系统、知识表示、人在回路中的人工智能 (human-in-the-loop AI)、搜索、规划、推理、机器人学和感知、伦理学以及结合 AI 的医疗、交通和商业等。

### 三、邀请报告

AAAI-21 会议一共邀请了 8 位杰出的演讲者，演讲内容主要包括深度学习黑盒理论、语言预训练模型、超人工智能，以及人工智能如何赋能医疗保健和传统生产行业；同时探讨了人工智能的广泛利用对个人隐私数据的影响。下面将介绍 8 位杰出演讲者的演讲内容。

**What Can and Should Humans Contribute to Superhuman AIs?** 超人人工智能不是指像人类一般的人工智能，而是使世界变得更美好的特定应用程序的人工智能。卡内基梅隆大学 Tuomas Sandholm 介绍了他在组合市场、器官交换和不完全信息游戏设置中的超人工智能领域的丰富经验。他讨论了如何发明和界定新的人工智能应用程序。在将政策优化和组合学留给人工智能的同时，人类应该如何提供价值框架将目的和方法分离开来，并以一种可扩展的方式对非常大规模的动态问题进行面向未来的优化。当定理（不仅仅是证明）和经验主义理论需要很长时间以至于超出人类理解的时候，Sandholm 对科学的未来感到疑惑。他讨论了什么是可解释性，以及为什么在许多人工智能应用中不应该要求它。最后他建议将伦理从事后讨论活动转变为一个系统设计学科，即预设计伦理。

**Recent Advances in Language Model Pretraining.** 近年来在大量原始文本上预先训练的大型语言模型彻底改变了自然语言处理。华盛顿大学 Luke Zettlemoyer 讨论了最近在语言模型预训练方面的工作，从 ELMo、GPT 到最近的序列到序列模型的预先训练方法，如 BART、mBART 和 MARGE，它们提供了迄今为止最普遍适用的方法。

**Opening the Black Box of Deep Learning (+ Takeaways for AI).** 深度学习的成功已经使一些经典的人工智能任务取得了显著进展，但不能实现完全人工智能。普林斯顿大学 Sanjeev Arora 认为目前关于深度学习的黑盒观点是实现完全人工智能的巨大障碍，并通过举例说明了他的观点。最后他认为利用数学理解黑盒现象对实现完全人工智能至关重要。

**Digital Learning Coming to Life.** 2020 年席卷

世界的全球大流行病迫使许多教学活动转为了数字形式，使用在线平台提供教学交流。这种转变既是挑战也是机会。数字学习比传统学习更具包容性，并且打开了关于教学和学习的的海量数据收集的大门，这将使我们能够开发和部署人为驱动或人工智能驱动的干预措施，从而进一步提高学生的参与度和成功率。

**How AI and Technology Design will Dictate Our Civic Future.** 随着技术的进步，采用哪种技术或不采用哪种技术很可能会重新定义每一种人口统计学价值和法律。从而使数据驱动的算法和技术设计师对我们的日常生活和治理国家的准则有很大影响，例如第一波挑战就是个人数据隐私。我们不知道这一切将如何融为一体或分崩离析？

**A Tale of Two Translations.** 麻省理工学院 Regina Barzilay 介绍了她的团队在癌症诊断和药物设计这两个医疗保健领域开发和部署机器学习方法的经验。她集中讨论了实现这些功能的算法，同时强调尚未解决的技术挑战。此外，她还分享了在此过程中获得的经验教训，最后对如何扩大对医疗保健的影响，向人工智能社区提出了建议。

**AI Infusion Investment Outlook.** 李开复博士讲述了人工智能如何利用深度学习突破从重大科学发现时代走向商业实施时代。他深入探究了商业社会对人工智能注入传统行业的新阶段的期待，并阐述了企业将如何采用人工智能来提高生产率，实现数字转型。

**Talking to the Public about AI.** 牛津大学 Michael Wooldridge 讲述了他在向非专业受众谈论人工智能领域进展，以及大众对于人工智能未来走向的真实期待。

### 四、会议优秀论文

AAAI-21 会议出现了三篇最佳论文、三篇最佳提名论文以及六篇杰出论文。涉及的内容有长序列时间序列预测、多智能体学习、因果推理、序列推荐、监督学习、无监督学习、强化学习、对抗攻击和增量学习等方面。

主 track 二篇最佳论文介绍如下。

**Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting.** 长序列时间序列预测(LSTF)要求模型能够有效地捕获输出和输入之间的精确远程依赖耦合关系。然而 Transformer 存在几个严重问题使其不能直接适用于 LSTF, 例如二次方时间复杂度、高内存使用量和编码器-解码器体系结构固有局限性。为了解决这些问题, 该论文设计了一个高效的基于 LSTF 模型, 命名为 Informer。在 4 个大规模数据集上的大量实验表明, Informer 方法在很多方面均显著优于现有方法, 为 LSTF 问题提供了一种新解决方案。

**Exploration-Exploitation in Multi-Agent Learning: Catastrophe Theory Meets Game Theory.** 为了在多智能体学习(MAL)的探索开发取得进展, 该论文研究了一种类似于 Q 学习的平滑方法, 此学习模型作为研究勘探开发的最佳模型具有很强的理论依据。该论文证明了其成本模型明确捕捉了博弈成本和探索成本之间的平衡, 并且在具有异质学习代理的加权潜在博弈中, 它总能收敛于有界理性下博弈的标准解概念——量化响应均衡(QRE)集。该论文提供了一个形式化的理论, 证明了可以通过调整探索参数进行平衡选择, 从而对系统性能产生积极影响或消极影响。

次 track 人工智能对社会的影响最佳论文介绍如下。

**Mitigating Political Bias in Language Models through Reinforced Calibration.** 当前的大规模语言模型可能会因为它们所基于的数据而产生政治偏见, 当它们被部署在现实世界中时, 可能会导致严重的问题。该论文描述了衡量 GPT-2 生成的政治偏见的指标, 并提出了一个强化学习框架来减轻生成文本中的政治偏见。通过使用来自单词嵌入或分类器的奖励,

在没有访问训练数据或需要重新训练模型的情况下, 强化学习框架指导生成偏见。在对政治偏见敏感的三个属性(性别、位置和话题)的实证实验中, 提出方法在保持可读性和语义一致性的同时, 根据指标和人类评价减少了政治偏见。

## 五、研讨会

AAAI-21 研讨会计划包括 26 个研讨会, 涵盖了人工智能的广泛主题, 分别为情感内容分析、用于行为改变的 AI、用于城市交通的 AI、人工智能安全、在紧急情况下打击区域语言中的在线敌对帖子、常识知识图谱、内容创作和设计、图深度学习的方法和应用、为远程医疗设计的 AI、第九届对话系统技术挑战赛、人工智能中的可解释代理、学习和推理中的图和复杂结构、第五届国际健康智能研讨会、混合人工智能、后冠状病毒时代的 AI 教育想象、金融服务中非结构化数据的知识发现、训练期间可学习的网络架构、元学习竞赛、用于计算机视觉的元学习、规划行动和意识识别(PAIR 2021)、保护隐私的人工智能、人机对话的推理和学习、游戏中的强化学习、科学文献理解、迈向鲁棒安全高效的机器学习、值得信赖的 AI 医疗。

## 六、总结与展望

受新冠疫情的持续影响, AAI-21 会议延续了完全线上虚拟会议的形式。无论是在 AAI-21 会议的投稿论文还是接收论文中, 机器学习、计算机视觉依旧高居榜首, 自然语言处理则较去年有所下降。随着人工智能理论不断发展, 如何将 AI 技术更好地应用于人类的生产生活成为不可忽视的问题。

责任编辑 崔海楠



王金甲

燕山大学信息科学与工程学院教授, 主要研究方向为信号处理和模式识别。担任中国计算机学会计算机视觉专委会委员。

Email: wjj@ysu.edu.cn

## 北京科技大学殷绪成教授访谈

2021年1月12日,《CCF-CV专委简报》在线采访了北京科技大学计算机与通信工程学院模式识别与人工智能技术创新实验室主任殷绪成教授。下面是采访实录。

殷老师,您好!首先,请您跟大家分享一下您的个人研究经历。

我2006年博士毕业于中国科学院自动化研究所,博士论文为“金融票据识别系统的应用研究”,针对金融票据(银行票据、保险保单等)文档图像分析与文字识别进行方法研究,并结合汉王OCR产品与系统进行了技术应用,率先在国内研制了大规模应用的银行票据识别系统。

2006年至2008年,在富士通研究开发中心担任研究员(Scientific Researcher),主要从事手机拍照文档图像分析与识别研究,发明了文档图像透视形变快速矫正技术,研发了业内首款手机文档拍照自动在线矫正处理应用程序(F905i/F906i/F01A,2007年)。

2008年6月,回到北京科技大学计算机系从事教学科研工作,从事模式识别、文字识别、计算机视觉研究,先后任副教授、教授,担任模式识别与人工智能技术创新实验室、北京科技大学-亿智电子科技人工智能联合实验室主任。十多年来,一直以大规模、强复杂自然场景/网络图片/复杂视频文本检测、跟踪与识别为中心,开展方法研究与技术创新,研制了网络图像文本识别技术超大规模应用系统。

您在文字识别及文档图像分析与识别领域取得了很大成就,能否分享一下您对这个领域的研究现状和未来发展情况的认识?针对这个领域的研究者,您有什么建议?

文字识别、文档图像分析与识别是模式识别、计算机视觉中的一个经典领域,也是整个人工智能中最早的一个热点研究领域。20世纪60年代初,IBM推出了世界上首个商业化OCR产品,也是世界上早期商业化人工智能产品的代表。80、90年代,文字识别相关论文占据了IEEE T-PAMI期刊的小半边江山。

目前,随着移动互联网、智能物联网、人工智能的迅猛发展,涌现了海量的自然场景图像、网络图像视频、历史文化文档、财务财会票据、法律法规文件等电子化文档图像数据,文字识别又一次成为模式识别中一个重要的热点研究领域。其中,自然场景、网络图像文本检测与识别,特别是多语言文本检测与识别,依然是当前及将来的热点研究。同时,个人认为,文字识别研究未来发展还包括以下几个问题:

(1) 小样本学习方法及可解释性问题,不仅是模式识别、机器学习、人工智能中的核心问题,更是多语言、跨场景文字识别研究中现实的基础问题;

(2) 自然语言处理技术及多模态信息融合问题,即嵌入自然语言处理技术、融合文本/图像/视频等多模态信息,服务于文档图像分析与识别,将是未来文字识别研究的一个趋势;

(3) 复杂中文手写文档图像分析与识别问题，一直都是文档图像分析与识别领域的一个难点问题，也是中国文字识别科研工作者当仁不让的责任，需要国内学术界、产业界大力支持与通力合作。

您作为第一完成人，负责的“网络图像视频大数据的智能识别关键技术及应用”项目获得了2019年度北京市科技进步一等奖，能否分享一下这项成果？您的感受或者“获奖感言”是什么？

我们的“网络图像视频大数据的智能识别关键技术及应用”项目有幸获得2019年度北京市科技进步一等奖，虽然本人是第一完成人，主要还是团队、合作伙伴大家一起努力的结果。

项目提出了鲁棒模型构建、快速增量匹配等新方法，突破了文字、目标、视频智能识别等关键技术，构建了先进的文字识别及多媒体大数据智能分析平台架构，建设了面向国家互联网信息实时分析的图像识别与信息管理系统、面向城市物联网的智能感知与市政城管应用系统等多个超大规模技术应用系统，有力地保障了国家网络信息规范利用和城市管理，产生了重要的社会效益和经济效益。

该项目是一个典型的产学研深度融合实例。自从2008年回到北京科技大学从事教学科研工作以来，本人及团队以应用研究为主，针对模式识别、文字识别、计算机视觉领域中的共性技术挑战，先后和富士通、三星、汉王、科大讯飞、腾讯等单位，进行了良好的技术交流与合作研究，产学研成效明显，不仅推进了人工智能技术创新与成果落地，也提升了学校人工智能人才培养质量。

您连续四届（2013、2015、2017和2019年）荣获国际文档分析与识别大会技术竞赛文本检测和文本识别等15项冠军，这非常难得，请问您是如何做到的？能传授一下您的经验么？

国际文档分析与识别大会 Robust Reading 技术竞

赛是国际模式识别领域代表性的重要经典赛事；从2003年设置以来，几乎世界上学术界、工业界所有重要的文字识别研究团队都参与了该项赛事，累计几百支团队参加了评测。

2013年是我们第一次参赛，当时我们提出了一种快速的文本检测新方法，比较幸运地赢得了那届比赛自然场景文本检测和网络图片文本检测双料冠军；同期，我们的新方法论文投稿 IEEE T-PAMI 也被接收了 (Robust Text Detection in Natural Scene Images, IEEE T-PAMI, 36(5): 970-983, 2014)；同时，当时的技术也做了手机 APP Demo，能够实时完成手机拍照场景文本检测，后来技术许可给了三星公司（2013年）。后面几届的竞赛越来越难，越来越激烈。2019年那届，上百支队伍参加了比赛，包括了国内最顶级的互联网、高科技及人工智能公司的研究团队。

个人觉得，参加技术竞赛，不管是国内比赛还是国际评测，最重要的是两点：

第一，对于竞赛主题要有一个深入的了解，最好有一个性能还不错的创新方法，能够解决竞赛主题的核心问题。这样，既可以通过竞赛验证方法的创新性和有效性，也可以通过竞赛激励自己，提升研究的兴趣。

第二，对于竞赛任务要有一个充分的认识，需要一个较长时间的准备工作，最好是一个研究小组一起准备。现在几乎每一项学术竞赛竞争都比较激烈，如果没有较长时间的应赛准备，而是匆忙参赛，一般很难取得理想的成绩，这样反而会影响研究的积极性。

您早期曾在汉王科技股份有限公司研发中心任研发工程师及技术经理，也曾在富士通研究开发中心信息技术部担任过研究员，请问您为什么最后还是选择了在高校工作呢？您觉得在公司和高校的最大区别是什么？

我先后在汉王科技、富士通研究开发中心工作了6年，主要从事文字识别、模式识别相关的研发工作。企

业工作的一个主要特点为，工作任务相对固定，工作进度相对严格，缺乏一定的自主性；如果希望从事一些自己感兴趣的研究课题，还是在高校更合适些，可以长期地进行探索，张弛有度。公司和高校另外一个区别就是，公司的研究工作具有很强的应用性和经济性，希望研究成果当年或尽快能够给公司带来直接的收益。而高校的研究工作具有长期性和公益性，侧重于原理方法的创新性和人才培养的有效性，当然，如果具有落地应用的价值则更好。个人认为，不管是在公司，还是在科研院所和高校做研究，方法或技术的创新性都是基本要求，都各有优势和局限性，“只有合适的，没有最好的”。

您担任模式识别与人工智能技术创新实验室主任，能否分享一下该实验室的宗旨、研究方向及其建设情况？

我们模式识别与人工智能技术创新实验室由 1 名教授、3 名副教授、2 名教师/博士后和 50 余名博士硕士研究生组成。实验室主要从事模式识别、文字识别、计算机视觉等领域的应用研究与技术创新，追求“顶天立地”，面向国家战略需求应用（互联网信息安全、人工智能芯片、智能制造）及人工智能垂直应用，重点研究创新性强、可用性高的共性关键技术。

2016 年，我们成立了北京科技大学-亿智电子科技大学人工智能联合实验室，专注于面向智能驾驶、智能安防、智能物联网等智能边缘计算应用的图像识别技术，重点关注人工智能芯片的算法研究与创新应用。经过几年的发展，在系统级（SoC）人工智能芯片技术研究与应用方面，取得了不错的进展，车牌识别、车辆识别、人脸识别等已在系统级人工智能芯片中进行软硬一体化设计与规模化应用。

您 2103、2014 和 2016 年 3 次在美国 University of Massachusetts Amherst 和 University of Massachusetts Medical School 进行了访问交流，能否分享一下您这段时间的主要研究工作及您对国外研究环境的看法？

我三次在 University of Massachusetts 进行访问交流，合作实验室包括智能信息检索中心、计算机视觉实验室和生物信息自然语言处理实验室，跨度比较大，印象比较深。

首先，从研究条件来说，中国跟美国等国外没有明显的差距，研究水平也互有优势，单从 CV、IR、NLP 顶会顶刊论文发表来看，在 2013、2014 年国内外就没有明显差别了。当然，在特定领域或特定点，我们还有很多事情需要继续努力。

同时，国内外高校教授对于具体研究工作的开展差别还是比较大的。国内很多高校里，往往强调大团队、大项目、大成果，知名教授很多精力花在团队管理、项目申报、成果宣传上，对于具体点、具体研究任务的长期跟进与深入探讨，则投入时间不多。在 University of Massachusetts 接触的几位教授，虽然他们也要花时间去申请项目，但是对于具体点的研究非常关注，对于具体方法和研究细节也比较投入。像 ACM Fellow Bruce Croft 教授，他是国际信息检索领域泰斗级人物，每周有固定几个上午和学生讨论课题与修改论文，而且非常认真、投入。这些给我留下了很深的印象。我也希望自己在关注团队、项目的同时，能够花更多的时间在具体研究内容及研究方法上，这也是我自己需要持续改进的地方。

您是如何将您的科研经验及成果融入教学之中的？

这是一个非常大的问题。高校工作的两个主要内容，教育教学和科学研究；但是，高校工作最本质的还是教育教学及人才培养。高校科研的核心要求之一就是，利用科学研究丰富教育教学方式，提升教育教学质量，推进人才培养。

在课程教学中，融入相关的技术方法和最新的科研成果，能够激发同学们的学习兴趣，提升学习效果。我

主讲《离散数学》、《软件工程》、《人工智能》等课程。在讲《离散数学》图论中加权图及最短路径问题时，除了介绍经典的 Dijkstra 算法，往往会扩展最短路径及动态规划方法，并介绍这些方法在物流路径规划、视频时序分析等领域的应用技术。这样，使学生感觉不仅仅在学这个具体的知识点，而是了解这个知识点的应用及其现实重要性。同时，我们还有设置一个 Course Project，利用最短路径方法来做一个北京地铁票价原型系统。

如果吐露研究工作者的心声，您最想说的是什么？

我在模式识别、文字识别、计算机视觉领域进行了近二十年的学习与研究，虽然没有取得很大的成绩，但是还是有一些不太成熟的心得，一些经验或者说是教训。其中，最重要的一条是，“做事情还是要专注。”我相信，在研究中，专注到一个具体的点，锲而不舍，坚持到底，数年后，就可能带来突破。

谢谢大家！

责任编辑 余焯 赵振兵



## 殷绪成

北京科技大学教授、博导，计算机与通信工程学院副院长，模式识别与人工智能技术创新实验室、北京科技大学-亿智电子科技人工智能联合实验室主任，中国图象图形学学会文档图像分析与识别专委会副主任/秘书长、中国自动化学会模式识别与机器智能专委会委员、中国计算机学会计算机视觉专委会委员、中国人工智能学会模式识别专委会委员。主要研究领域包括模式识别、文字识别、计算机视觉及人工智能芯片技术，近五年在中国计算机学会推荐的国际期刊和会议上发表论文四十多篇，连续四届（2013、2015、2017 和 2019 年）荣获国际文档分析与识别大会技术竞赛文本检测和文本识别等 15 项冠军，获 2019 年度北京市科技进步一等奖（第一完成人）、2018 年度教育部科技进步二等奖（第一完成人）。

## 委员好消息

❖ 2020年12月17日《人民日报》综合版整版刊登“典赞·2020 科普中国”宣传推选活动，宣传评选出来10名基层科普人物、10名科研科普人物、7名科普特别人物以及10个科普图书图文、10个科普影音视频、10个科普展览展品。CCF-CV专委会委员、复旦大学张军平教授在科学网等媒体连载的系列文章《爱犯错的智能体》被评为年度科普图书图文。

❖ 2020年12月31日，中国计算机学会公布了2020年“CCF 优秀博士学位论文奖”评选结果，10位博士获得2020年“CCF 优秀博士学位论文奖”，9位博士获得2020年“CCF 优秀博士学位论文奖”提名。CCF-专委会委员、中科院计算所陈熙霖研究员指导刘昊淼完成的论文《面向物体语义理解的视觉表示学习》、北京大学彭宇新教授指导何相腾完成的论文《辨识性特征学习及在细粒度分析中的应用》获得2020年“CCF 优秀博士学位论文奖”。CCF 优秀博士学位论文奖授予在计算机科学与技术及其相关领域的基础理论或应用基础研究方面有重要突破，或在关键技术和应用技术方面有重要创新的中国计算机领域博士学位论文的作者。

❖ 2021年1月10日，中国计算机学会2020年度优秀专委会评审结果揭晓，CCF-CV专委会荣获2020年度优秀专委会。

❖ 2021年1月13日，中国计算机学会公布了2020年CCF会士评选结果，本年度共10位CCF杰出会员当选，CCF-CV专委会委员、重庆邮电大学高新波教授因在异质图像合成与图象质量评价算法研究取得丰硕

成果、所开发的人脸画像合成与识别系统为维护社会公共安全提供技术支撑、长期服务学会发展、多次在CCF主办的重要学术会议上作报告而成功当选CCF会士。

❖ 2021年1月13日，北京市科学技术委员会公示了2020年度北京市科学技术奖初审通过项目，由CCF-CV专委会委员、中科院计算机蒋树强研究员和中国科学院大学黄庆明教授等完成的“多模态环境感知及适配交互技术与应用”拟授科技进步一等奖，由CCF-CV专委会委员、中科院计算所山世光研究员和阚美娜副研究员等参与完成的“大规模人工智能数据柔性生产关键技术及应用”拟授科技进步二等奖。

❖ 2021年1月17日，中国计算机学会公布了2020年度CCF杰出演讲者名单，本年度共评出47位杰出演讲者，CCF-CV专委会3位委员当选，他们是：厦门大学纪荣嵘教授、北京大学林宙辰教授和清华大学鲁继文副教授。

❖ 2021年1月15日，广东省科技厅公示了2020年度广东省科学技术奖拟奖项目名单，由CCF-CV专委会委员、中山大学郑伟诗教授、王昌栋副教授、赖剑煌教授、谢晓华副教授等完成的小样本视觉表征与识别方法拟授自然科学二等奖。

❖ 2021年2月26日，中国图象图形学学会发布了2020-2022年度中国图象图形学学会青年人才托举工程项目终评结果公告，共有4名入选者，CCF-CV专委会副秘书长、中科院自动化所黄岩副研究员入选。

责任编辑 刘海波

# 多器官分割开源代码

西安电子科技大学

张家铭 张亮

**器**官分割是医学图像领域中一项非常重要的任务，同时也是计算机辅助诊断和放射诊疗计划等临床应用的基础。由于器官类别间标注不平衡且不同器官间结构差异大等问题，多器官分割相比于常见的单器官分割是一项更具挑战性的任务。

本文主要介绍一份使用 PyTorch 框架实现的多器官分割开源代码，其中模型的基础结构为两个级联的 3D V-Net 网络。之后本文对其中的数据预处理、损失计算、网络训练和测试过程进行了详细分析。

## 1、数据处理

数据处理是影响深度学习模型性能的重要因素之一。代码中主要进行了层厚统一、分辨率减半、阈值截断和归一化四个数据处理操作。

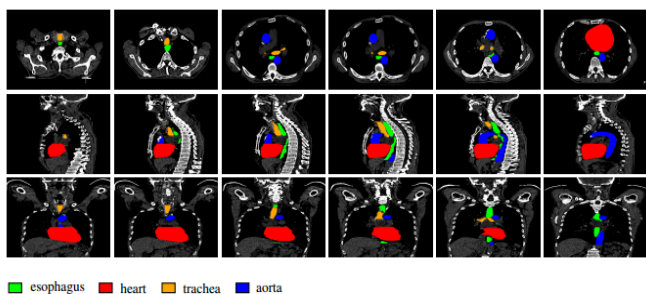


图 1 器官可视化

- (1) 层厚统一: 医学影像不同 CT 序列的层厚并不一致，因此在构建三维数据时需对所有 CT 序列的层厚进行统一。
- (2) 分辨率减半: 将分辨率减半后网络可以一次性输入更多的 CT 切片，在 3D 网络中表现更好。

(3) 阈值截断: CT 影像中人体腹部器官的 HU 值范围大部分都可以被包括在[-350, 350]之间。因此，将 CT 影像的 HU 值截断在此区间内可以有效排除无关区域的干扰。

(4) 归一化: 经过卷积运算后，特征中的体素值会被进一步放大，将 HU 值的范围进一步限制到[-1, 1]之间，更有利于网络的优化。

## 2、损失计算

多器官分割任务中由于标签不平衡，直接使用交叉熵损失函数容易导致小器官在损失计算过程中被大器官的值所“淹没”，因此代码中采用了 Dice Loss 进行损失计算。对于每一个 V-Net 网络，采用式(1)中的公式进行计算：

$$\mathcal{L} = 1 - \frac{1}{K} \sum_{k=1}^K \frac{2 \sum_i P_k(i) L_k(i)}{\sum_i P_k^2(i) + \sum_i L_k^2(i)}$$

其中  $P_k(i)$  和  $L_k(i)$  分别表示将第  $i$  个体素预测为类别  $k$  的概率和标签将其标注为类别  $k$  的概率(0 或者 1)。式(1)将预测概率图与标签概率图交集的两倍与其并集的比值作为单个类别的 Dice 系数，然后将所有类别的 Dice 系数相加再求平均值，得到该 V-Net 网络整体的 Dice 系数，1 减 Dice 系数后即可得到单个网络上的 Dice 损失。最终整个网络的损失为所有单个网络 Dice 损失的平均值。

## 3、网络训练与测试

3D 网络在训练与测试过程中需要的显存远远大于

2D 网络, 因此通常不将整个 CT 序列送入 3D 网络进行训练, 而是将 CT 序列拆成一个个由多张切片组成的数据块。

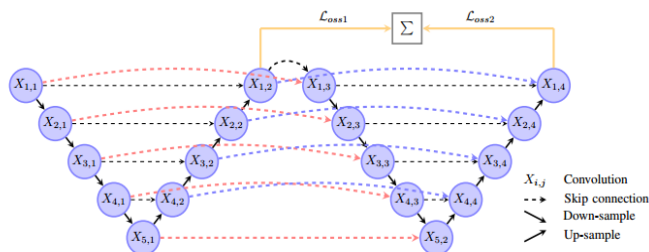


图 2 网络结构

**训练:** 首先根据标签确定器官的位置, 然后在包含器官中的切片中依次挑选 48 张连续切片送入网络。其中相邻的两次挑选的切片随机重合 0 到 24 张切片, 增加数据的多样性, 避免数据量少造成网络过拟合。整个网络采用深监督的方式进行训练, 对两个 V-Net 网络分别计算 Dice Loss, 之后求和构成最终的 Loss。

**测试:** 同样, 在测试的 CT 序列中挑选多个 48 张切片组成的数据块, 送入网络依次进行预测。随后拼接所有预测结果并计算 Dice 分数, 其中两个数据块之间重合的切片采取后一个数据块中的值。

#### 4、参数配置

代码中提供了参数的配置文件, 研究人员可以通过设置不同的参数微调网络或者应用于不同数据集中。主

表 1 代码参数说明

参数	意义	参数	意义
Model Name	模型名称	Slice Expand	选取切片构成数据块的挑选范围
Slice Size	数据块包含的切片数目	HU Upper	阈值截断时的 HU 值上限
Num Organ	数据集中器官的数目	HU Lower	阈值截断时的 HU 值下限
Loss Func	损失函数类型	Slice Thickness	统一层厚中的标准层厚
Down Scale	降低分辨率时的下采样倍率	GN	组卷积中的通道数

责任编辑 李策

要的参数和对应意义如表 1 所示。

#### 5、结果展示

训练得到的网络模型对验证集进行预测后, 先计算所有器官上的平均 Dice 分数并保存到一个 Excel 文件中。随后, 对模型预测的分割结果以及预测结果中错误的体素进行可视化。图 3 中, 子图(a)、(b)、(c)依次展示了 Baseline、BLSC、Mixed 三个模型的器官预测结果, 而子图(d)、(e)、(f)则分别展示了三个模型预测错误的像素。

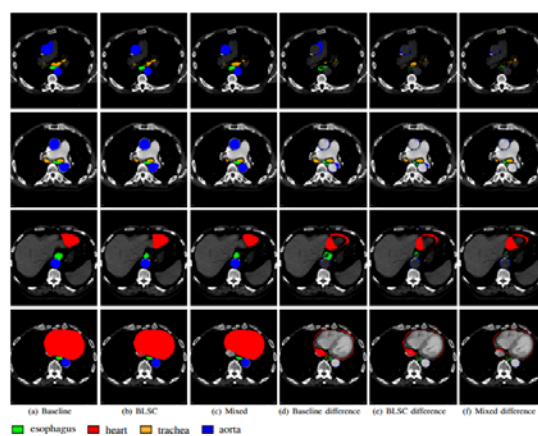


图 3 MADNet 不同算法下的网络结构和深度图效果

**论文:** <https://ieeexplore.ieee.org/document/9006924>

**代码:** <https://github.com/SSurprising/BLSC>

## 张家铭

硕士研究生, 西安电子科技大学计算机科学与技术学院, 主要研究方向为医学图像处理。

## 张亮

教授, 西安电子科技大学计算机科学与技术学院, 嵌入式技术与视觉处理研究中心主任。重点关注机器人复杂场景中环境感知与场景理解、深度学习、三维点云场景描述等热点方向。

个人主页: <https://web.xidian.edu.cn/zhangliang/>

# 基于 RGB-D 的显著物体检测数据集

大连理工大学 付陈平 樊鑫

**显著物体检测**(Salient Object Detection, SOD)是主体目标像素级检测的特殊检测方向，其主要目的在于找到图像/视频中最明显、最显著的物体，并将其从背景中精细地分割出来，可应用于图像/视频编辑、合成、分析等领域。基于人类视觉注意的显著物体检测是图像/视频处理的重要步骤。注意是人类视觉信息加工过程中的一项重要心理调节机制，它能够帮助我们分配有限的信息加工资源，使视觉感知功能具备选择能力。显著物体检测正是模拟人类视觉信息注意机制所提出并发展起来的。概括来说，显著物体检测的发展主要经历三个阶段。Itti 算法是最早的一批经典显著检测模型，它们的提出掀起了跨认知心理学、神经科学和计算机视觉等多学科多领域研究的第一波热潮。接下来，有部分研究人员将显著物体检测定义为二元分割问题，这推动了该方向的第二波研究热潮。深度学习的发展，特别是完全卷积神经网络的引入，拉开了显著检测的第三波发展序幕。与同时期的经典方法相比，基于卷积神经网络的方法打破了对手工特征的依赖，从而减轻了对中心偏见知识的需求，成为显著检测方法研究的关注点。此外，卷积神经网络模型先天具备大量可调参数和神经元，能够提供丰富的全局信息，可更好识别图像中的显著目标区域。卷积神经网络能够实现其他方法手段难以实现的性能，使其逐渐成为显著物体检测的研究主流。

近年来，RGB-D 已被广泛地应用于显著物体检测，随着如何拍摄高质量照片成为手机制造商间重要的竞争点，如何充分使用 RGB-D 信息进行显著物体检测已成为当前的研究焦点。面向研究发展需求，最近几年 SOD 领域出现多个 RGB-D 数据集。本文将详细介绍三

个代表性的 RGB-D SOD 任务数据集，包括 SIP、NLPR 和 STERE 数据集。

## 1、SIP 数据集

**介绍：**SIP (Salient Person, SIP) 数据集发布于 2020 年，由南开大学程明明团队收集发布。SIP 数据集是第一个以人类活动行为为主的 RGB-D SOD 数据集。该数据集共包括八个不同背景场景下的 929 张 RGB-D 数据图像，每幅图像中人物穿着不同，目标存在不同程度遮挡且具有复杂的形状变化。图 1 展示了 SIP 数据集中图像的例子。



图 1 SIP 数据集示例

如图 1 所示, SIP 数据集中具有不同数量的显著物体, 物体尺寸, 物体位置, 场景复杂度和照明条件的图像。此外, 该数据集同时提供了深度图和标注, 因此该数据集为研究人员提供了一个新的研究方向, 例如根据“RGB”和“灰色”图像进行深度估计以及实例级的 RGB-D 显著物体检测任务。该数据集使用华为 Mate 10 采集图像。Mate 10 的后置摄像头采用了徕卡 SUMMILUX-H 镜头, 光圈为 f/1.6, 并结合了 12MP RGB 和 20MP 单色 (灰度) 传感器。深度图是由 Mate 10 自动估计的。程明明团队请了 9 个人, 他们身着不同颜色的衣服, 在真实的日常场景中表演特定的动作。以此完成数据集的采集工作。程明明团队共收集 5269 张图像以及深度图, 此后, 团队手动挑选大约 2500 张图像, 保证每张图像都包含一个或多个显著人物。经过认真标注后, 又进一步筛选出前 1000 张令他们团队最为满意的图像。最后, 对低质量标注图像进行下一轮丢弃, 从而获得 929 张高质量标注的真值图像。该数据集进一步从中心偏向、物体大小、背景物体、物体边界条件和显著物体数量这 5 个方面对数据集进行了较为全面的数据集统计分析。更多有关该数据集的详情可参考发布该数据集的论文“Rethinking RGB-D Salient Object Detection: Models, Data Sets, and Large-Scale Benchmarks”, 数据集相关介绍论文地址: <https://ieeexplore.ieee.org/document/9107477>。

**数据集地址:** <http://dpfan.net/SIPDataset/>

## 2、NLPR 数据集

**介绍:** NLPR 数据集发布于 2014 年, 由中国科学院和厦门大学共同联合组建。2014 年前, RGB-D SOD 任务极大受限于小规模和低分辨率的数据集, NLPR 的发布在很大程度上克服了这一障碍从而促进了 RGB-D SOD 任务的发展。该数据集是一个大规模的 RGB-D 数据集, 图像分辨率为 640\*480。数据集组建团队从不同场景中捕获了 5000 幅图像及其对应的深度图。在预处理和注释之后, 从它们中挑选 1000 张组成了最终的基准数据集 NLPR。图 2 展示了 NLPR 数据集中图像的例子。接下来, 本文将较为详细地介绍该数据集

的相关内容。

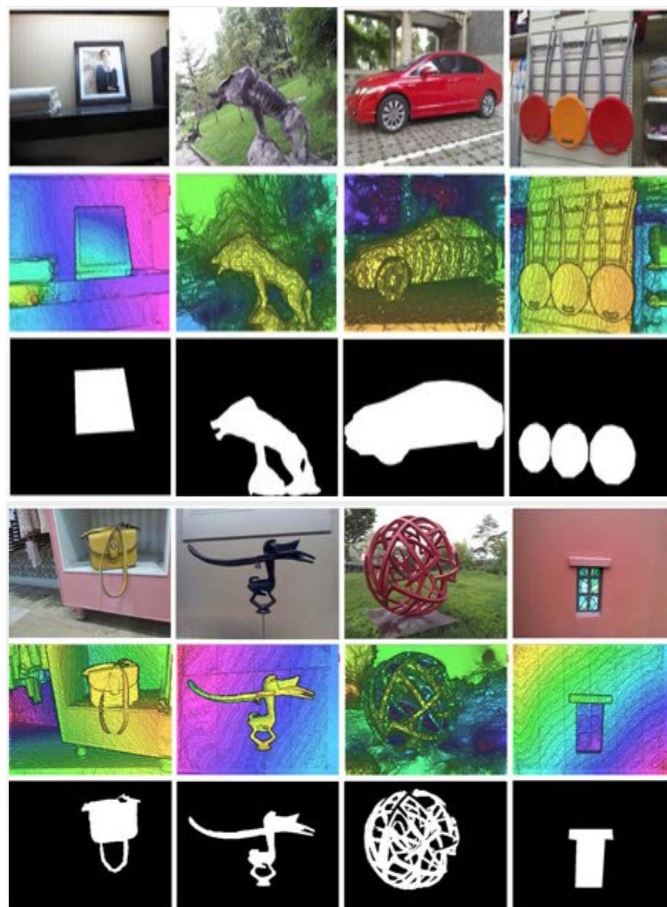


图 2 NLPR 数据集示例

NLPR 数据集使用标准的微软 Kinect 构建参考深度图。为避免相机抖动和拍摄数据模糊, 数据采集团队将 kinect 绑在一个坚固的三脚架上, 采集到的运动输出数据由相连的笔记本进行同步记录。团队使用改造后的 Kinect 设备捕捉室内和室外的一系列地点, 比如办公室, 超市, 校园, 街道等。具体来说, 总在阴天或阳光充足的黄昏对室外场景进行捕捉, 以避免阳光直射, 影响红外深度相机的精度。此外, 为了减少人类偏好造成的采集偏差, 每个场景都由一对收集器拍摄, 且每个物体至少从深度 0.5 到 10 米不等的四个方向进行拍摄。该团队共收集 5000 张自然图像及其深度图, 他们首先手动选择 2000 张图像, 保证每张图像都包含一个或多个独特的前景物体。然后, 对于每一张选定的图像, 要求五名参与者根据他们第一眼看到最吸引人注意力的物体画一个矩形。由于不同的人对同一幅图像中的显著对象可能会有不同看法, 因此他们进一步删除了标记一致性

较低的图像, 根据此做法选择了前 1000 张满意的图像。最后, 对每张图像中的显著目标进行手动分割。该数据集进一步从中心偏向、多样性、复杂性这 3 个方面对数据集进行了较为全面的数据集统计分析。更多有关该数据集的详细情况可参考发布该数据集的论文“RGBD Salient Object Detection: A Benchmark and Algorithms”, 数据集相关介绍论文地址: <https://ieeexplore.ieee.org/document/9107477>。

**数据集地址:** <https://sites.google.com/site/rgbdsalient/dataset>

### 3、STERE 数据集

**介绍:** STERE 数据集是 RGB-D SOD 领域第一个立体图像集, 于 2012 年由美国波特兰州立大学和山东大学联合发布。该数据集包括 1000 张立体图像, 图 3 展示了该数据集中的图像例子。

数据集采集团队首先从三个图像网站 Flickr, Stereoscopic Image Gallery 和 NVIDIA 3D Vision Live 下载收集 1250 张图片(图像分辨率为 [251~1200]\*[222~900]), 这些网站的图像往往由立体相机拍摄而来。接下来由 3 位工作人员同时标注同一张图像中的显著目标。此后那些删除标签最不一致的图像。具体来说, 他们根据标记的矩形间的重叠程度对图像进

行排序, 最终选取前 1000 张图像。该数据集的每张图像中最多有一个显著目标, 标注质量较高, 且该数据集具有较高的中心偏置。更多有关该数据集的详细情况可参考发布该数据集的论文“Leveraging Stereopsis for Saliency Analysis”, 数据集相关介绍论文地址: <http://web.cecs.pdx.edu/~fliu/papers/cvpr2012.pdf>。

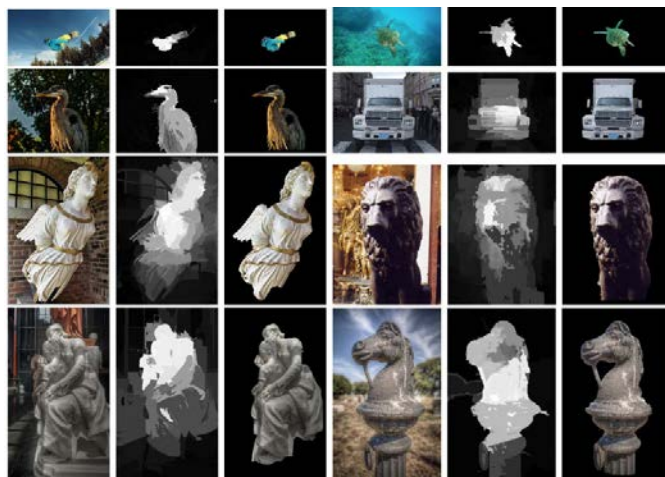


图 3 STERE 数据集示例

#### 数据集地址

<http://web.cecs.pdx.edu/~fliu/project/stereo-saliency/>

责任编辑 沈沛意



#### 付陈平

博士研究生, 大连理工大学国际信息与软件学院, 研究方向为计算机视觉。



#### 樊鑫

博士生导师, 大连理工大学国际信息与软件学院从事教学与科研工作, 担任中日国际信息与软件学院院长。研究方向为计算机视觉与图像处理、医学影像分析。

个人主页: [http://faculty.dlut.edu.cn/Xin\\_Fan/zh\\_CN/index.htm](http://faculty.dlut.edu.cn/Xin_Fan/zh_CN/index.htm)

## 好文推荐

悉尼科技大学团队“用不确定性修正域自适应中的伪标签”最新成果发表在 IJCV 2021。

论文：Zhedong Zheng, Yi Yang. Rectifying Pseudo Label Learning via Uncertainty Estimation for Domain Adaptive Semantic Segmentation, IJCV, Jan, 2021.

本文研究的是领域迁移问题中错误的伪标签问题，探讨了如何自动设定阈值来修正这种伪标签学习。

现有算法的伪标签往往通过人为设定阈值的方式来学习高置信度的伪标签，而忽略低置信度的标签。但是这个阈值往往很难选取。我们在思考如何设定这个阈值时，发现这个阈值取决于源域和目标域的相似程度。如果源域和目标域完全一致，那么阈值可以设为 0.999，大部分数据都有很确定的标签，全部的样本都可以拿来训练。但如果源域和目标域完全不一样，那么无论阈值设置为多少，所有的伪标签都是含噪声的。这种人为设定阈值的方式是不合理的。因此，本文提出用不确定性作为阈值，即自动学习阈值。

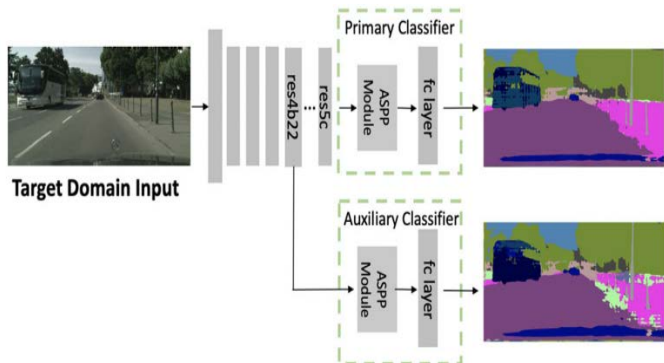


图 1 基于 Deeplab-v2 的双分类器模型

本质思想很简单，利用分割模型中的辅助分类器，如图 1 所示。一个主分类器连接在 res5c 后，一个辅助分类器连接在 res4b22。而伪标签发生错误的地方，往

往是两个分类器预测结果不同的地方，如图 2 所示。

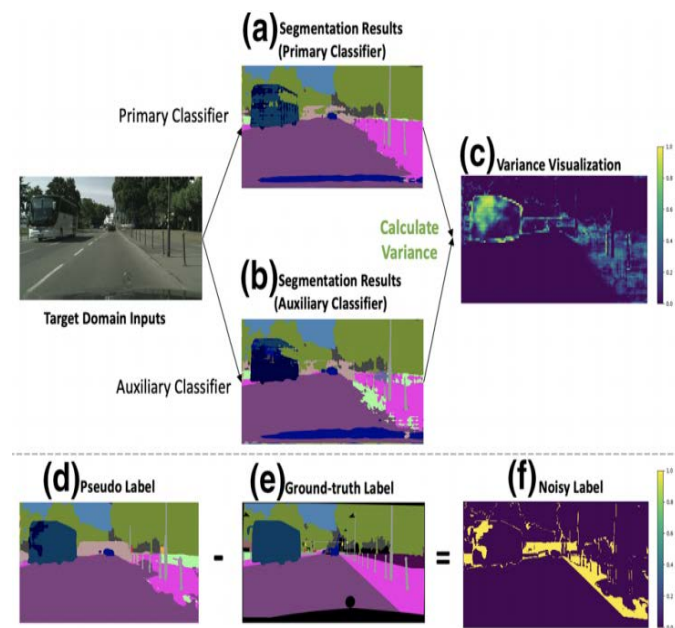


图 2 两个分类器间的预测差异

因此，我们对交叉熵损失进行了调整，公式如(1)和(2)所示。其中  $D_{kl}$  就是主分类器和辅助分类器预测结果的 KL 距离，如果差异大，则这个距离也就大，那么  $L_{rect}$  对于这种不确定的样本，就不惩罚(因为伪标签很可能是错的)。如果没有后面  $+D_{kl}$  这一项，模型会趋向把所有伪标签都看成是不确定的，那么  $L_{rect}$  就等于 0 了。为了避免这种情况，所以我们加了一个  $+D_{kl}$ 。

$$L_{rect} = E \left[ \exp \{ -D_{kl} \} L_{ce} + D_{kl} \right] \quad (1)$$

$$D_{kl} = E \left[ F(x_t^j | \theta_t) \log \left( \frac{F(x_t^j | \theta_t)}{F_{aux}(x_t^j | \theta_t)} \right) \right] \quad (2)$$

所提算法在不仅在常用数据集中优于传统算法，而且能够准确分割出图像中所占面积较小的目标。

责任编辑 贾同 樊鑫

## 好文推荐

浙江大学、商汤科技和南方科技大学团队的“基于形状先验引导实例视差评估的立体三维目标检测”最新成果发表在 IEEE/CVF CVPR 2020。

论文: Jiaming Sun, Linghao Chen, Yiming Xie, Siyu Zhang, Qinhong Jiang, Xiaowei Zhou, Hujun Bao. Disp R-CNN: Stereo 3D Object Detection via Shape Prior Guided Instance Disparity Estimation, IEEE/CVF CVPR, pp. 10548-10557, 2020.

随着深度学习热潮地掀起，越来越多的计算机视觉任务开始使用深度学习方法进行解决。基于深度学习的方法利用卷积神经网络解决了立体对应匹配问题取得了令人印象深刻的成果。三维物体检测方法用估计出的视差图作为输入，将其转换为深度图或点云，检测其中的物体。然而，由于视差估计网络是为一般的立体匹配而设计的，而不是用于三维目标检测任务，这些通道存在两个主要缺陷。首先，视差估计过程是在全图像上进行的，通常无法在低纹理或非朗伯表面上产生精确的视差，而这些正是成功进行三维边界盒估计所需的区域。此外，由于前景中感兴趣的物体通常比背景占据的空间小得多，视差估计网络和三维检测器在不需要检测物体

的区域上花费了大量的计算，导致运行速度较慢。

为此，浙江大学、商汤科技和南方科技大学团队提出了一个为实例级视差估计而设计的立体网络 Disp R-CNN 来检测 3D 对象，如图 1 所示。网络训练的视差估计仅在包含感兴趣对象的区域上执行，从而使网络聚焦于前景对象并学习适合于 3D 目标检测的类别特定形状先验。这样的算法减少了视差估计过程中输入和输出像素的数量，减少了代价量搜索的范围，从而减少了整个三维检测流水线的运行时间。为了解决当前缺乏像素级的真实基础注释的问题，该团队提出了一个伪地面真值生成过程，通过物体形状重建和渲染来获得准确的实例视差和实例分割掩码。在多种几何约束下，利用基于 PCA 的统计形状模型重构对象网格，能够在没有激光雷达的情况下为网络训练提供密集的监督。

该团队在 KITTI 上对模型进行评估。实验结果表明，在物体形状的先验指导下，估计的实例差异捕获平滑的形状和物体边界的尖锐边缘，同时比全帧的对应对象更精确。因此，与依赖全帧差异的最先进的基线 3D 检测器相比，3D 物体检测性能可以大大提高。

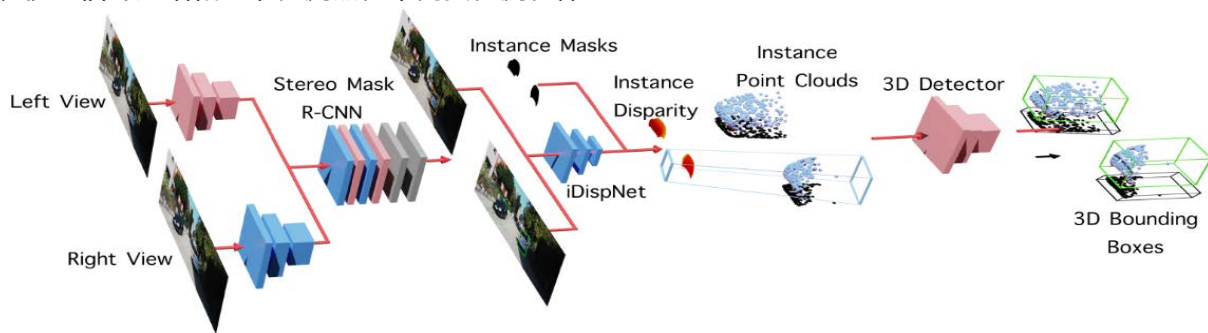


图 1 DECAMEL 算法流程图

责任编辑 贾同 樊鑫

## 好文推荐

清华大学、澳大利亚国立大学、悉尼科技大学和德州大学圣安东尼奥分校团队联合研究的“基于部分级卷积特征的行人重识别”最新成果发表在 IEEE TPAMI 2021。

论文: Yifan Sun, Liang Zheng, Yali Li, Yi Yang, Qi Tian, Shengjin Wang. Learning Part-based Convolutional Features for Person Re-identification, IEEE TPAMI, vol. 43, no. 3, pp. 902-917, Mar. 2021.

在针对行人重识别的算法研究中发现, 提取 part-level 的特征可以为行人图像的描述提供细粒度的信息。因此, 本文提出了一种基于 part-level 卷积特征的行人重识别算法, 主要贡献如下:

1) 提出了一个通用的 part-level 特征学习方法——Part-based Convolutional Baseline (PCB), 如图 1 所示。在 PCB 模块中, 输入图像通过主干网络中的层叠卷积层向前移动, 形成三维张量  $T$ 。PCB 提取  $T$  上的几个 part, 然后将同一 part 中的列向量平均为一个列向量  $g$ , 卷积核尺寸为  $1 \times 1$  的卷积层将  $g$  的  $p$  个模块转换为降维向量  $h$  的  $p$  个模块。最后, 将每个列向量  $h$  分别输入分类器中, 每个分类器由一个全连接层和 softmax 函数构成。在测试过程中,  $g$  或  $h$  的  $p$  个模块连接起来形成图像最终的描述特征。给定一个输入图像, PCB 得到由多个 part-level 特征组成的卷积描述符, 它包含多

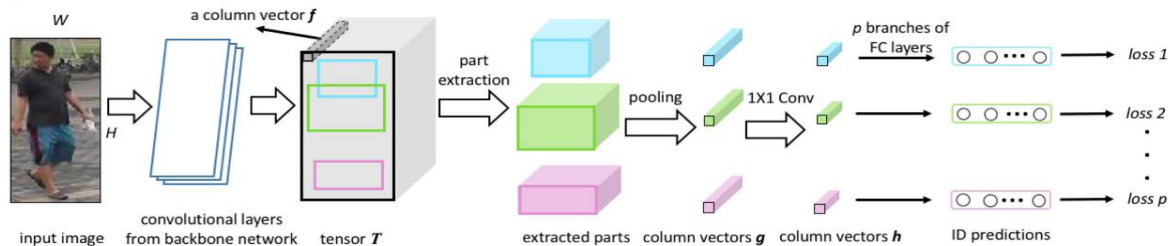


图 1 PCB 框架

种 part 划分策略, 包括姿态估计、人体解析和统一 part 划分。

2) 在 PCB 的基础上, 作者提出了 Refined Part Pooling (RPP), 重新调整每个 part 中的异常值, 让 parts 定位更加精准(防止分 part 时属于同一 part 的某些部位分到的另一个 part 中), 如图 2 所示。作者用一个矩形而不是立方体来表示三维张量  $T$ , 是考虑到所提算法更关注空间分割,  $T$  之前的层与图 1 保持一致, part 分类器预测每个列向量属于  $p$  模块的概率。然后, 以相应的概率作为采样权值, 从所有列向量中对每个 part 进行采样。GAP 表示全局平均池化。

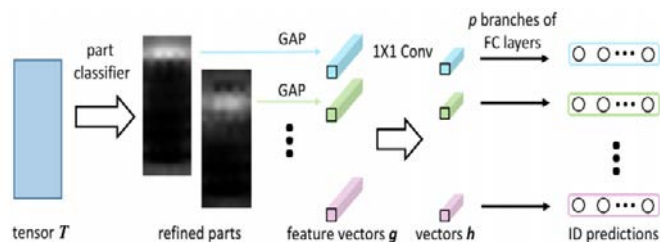


图 2 PCB 与 RPP 的结合框架

所提算法将 PCB 和 RPP 结合起来, 可以得到可靠的行人重识别结果。实验证明, RPP 可以进一步提升 PCB 的性能, 如: 在 Market-1501 数据集上, 所提算法的 mAP 为  $(77.4+4.2)\%$ , rank-1 精度达到了  $(92.3+1.5)\%$ 。

责任编辑 樊鑫 贾同

## 征文通知

### 1 会议征文

计算机视觉领域相关国内外会议的征文通知如表 1 所示。同时，可继续关注每个会议举办的 workshop 或 special session。

### 2 期刊征文

计算机视觉领域近期相关期刊专刊的征文通知如表 2 所示，包括 IEEE Journal of Selected Topics in Signal Processing, Image and Vision Computing, Pattern Recognition Letters 和 Image and Vision Computing。

### 3 会议简介

中国模式识别与计算机视觉学术会议 PRCV (Chinese Conference on Pattern Recognition and

Computer Vision)，由中国人工智能学会 (CAAI)、中国计算机学会 (CCF)、中国自动化学会 (CAA) 和中国图象图形学学会 (CSIG) 联合主办，定位国内顶级的模式识别和计算机视觉领域学术盛会。

第四届 PRCV 将于 2021 年 10 月 29 日至 11 月 1 日在北京国际会议中心举行，由北京科技大学、北京交通大学和北京邮电大学共同承办，中山大学、清华大学协办。本届会议将主要汇聚国内从事 PRCV 理论与应用研究的广大科研工作者及工业界同仁，共同分享我国 PRCV 领域的最新理论和技术成果，为大家提供精彩的学术盛宴。现向广大科技工作者公开征集高质量、原创性的优秀论文。会议论文集将由 Springer 出版社 LNCS 系列出版，并被 EI 和 CPCI-S 检索。

责任编辑：刘帅奇

表 1 计算机视觉领域相关国内外会议

会议名称	会议时间	会议地点	截稿日期	会议网站
FG 2021	2021.12.15-18	Jodhpur, India	2021.04.15	<a href="http://iab-rubric.org/fg2021/">http://iab-rubric.org/fg2021/</a>
ACM MM 2021	2021.10.20-24	Chengdu, China	2021.04.17	<a href="https://2021.acmmm.org/">https://2021.acmmm.org/</a>
NeurIPS 2021	2021.12.13-17	Bilbao, Spain	2021.05.27	<a href="https://2021.ecmlpkdd.org/">https://2021.ecmlpkdd.org/</a>
ICMI 2021	2021.10.18-22	Montreal, Canada	2021.05.27	<a href="https://icmi.acm.org/2021/">https://icmi.acm.org/2021/</a>

表 2 计算机视觉领域相关国内外期刊专刊

期刊名称	专刊题目	投稿网址	截稿日期
IEEE JSTSP	Signal Analysis for Detection and Monitoring of Contagious Diseases	<a href="https://signalprocessingsociety.org/blog/ieee-jstsp-special-issue-signal-analysis-detection-and-monitoring-contagious-diseases">https://signalprocessingsociety.org/blog/ieee-jstsp-special-issue-signal-analysis-detection-and-monitoring-contagious-diseases</a>	2021.05.30
PRL	Application of Pattern Recognition in Digital world: Security, Privacy and Reliability (APRDW)	<a href="https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/application-of-pattern-recognition-in-digital-world">https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/application-of-pattern-recognition-in-digital-world</a>	2021.06.20
PRL	Few-shot Learning for Human-machine Interactions (FSL-HMI)	<a href="https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/">https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/</a>	2021.07.20
IVC	Deep Learning for Panoramic Vision on Mobile Devices	<a href="https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/">https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/</a>	2021.08.01

## 心底无私视界宽 ∞ 袁保宗教授专访

2020年12月17日,嫦娥五号返回器历经23天,成功在内蒙古四子王旗预定区域着陆,圆满完成了我国首次月球采样返回任务。这一成功的背后,蕴含了大量计算机视觉技术的应用。实际上,计算机视觉最早的应用之一也与探月有关。在1964年美国的月球探测飞船Ranger 7在探测月球时,为保证拍摄的视频不产生畸变,就在摄像头前增加了十字形的栅格,并通过计算机视觉的算法来实现纠偏。随后,计算机视觉的应用领域变得越来越广泛,新颖的算法也层出不穷。而每一个算法的提出,都饱含着计算机视觉领域科研工作者的心血。

自50年代以来,我国就已经在计算机视觉领域展开了相关的科研工作。而今,我国已经拥有了一支庞大的、在这一领域辛勤耕耘且能与世界一流水平并驾齐驱的科研队伍。在这一过程中,有一批见证了视觉领域的



图1 专访合影。北京交通大学袁保宗教授中间就座

发展,为我国计算机视觉领域的奠基做出了重大贡献的先驱。2020年12月14日,中国计算机学会计算机视觉专委会(CCF-CV)《CCF-CV专委简报》有幸拜访了北京交通大学袁保宗教授。本次由明悦为主采访和整理实录内容。

**为能更好地帮助我们回顾,以下我们采用了问答加书面回顾的形式来表述,其中回顾部分用粉色字体标识。**

**明悦(后缩写为明):** 我们想先来了解一下您在过去是如何踏上学术的这条道路,然后怎样展开相关的学术研究,我们了解到,您曾留学苏联,在列宁格勒工程学院获得了技术科学的副博士学位。您能否跟我们介绍一下您当时是怎样去留学的,在那边学的是什么专业,这一段过程对您的成长以及后面工作发展的影响,包括对我国信号处理以及计算机视觉领域作出杰出贡献有哪些帮助?您如果有难忘的人和事,也希望您跟我们分享一下。

袁保宗教授是信号与信息处理专家,资深教授。1950年-1953年,就读于北方交通大学。1960年获苏联列宁格勒铁道工程学院技术科学副博士学位。曾任北京交通大学教授及信息科学研究所首任所长,电子工程与信息技术学院院长。担任科技部国家重点基础技术研究项目“973”信息领域咨询专家,国家自然科学基金通信信息领域评审专家,国务院学位委员会第二、三、四届评议组成员,中国电子学会和中国通信学会常务理事,中国电子学会信号处理分会主任委员等,1986年获全国五一劳动奖章等各种奖项。1992年英国国际电机工程学会(IEE会士;北京分会副主席)。2000年获IEEE国际千年奖(注:国内一共近10位)和奖章。2018年11月24日,获2018年度CCF-CV终身学术贡献奖。

**袁保宗教授：**我就你之前发过来的采访提纲，做了一些修改和归纳总结，如下。

### 1、当初为什么要选择这个研究方向？

袁保宗教授长期致力于语音、图像、计算机听觉、视觉、多媒体通信与信息处理方面教学和科学研究。先后主持完成“智能视听信息处理系统”、“交互式会议电视系统”等多项国家级和铁道部重点课题科研项目，有多项创新及开拓。参与创建中国电子学会信号处理分会，积极开展国内外学术交流。为信息科学领域培养了大批高级人才。为我国信息技术发展做出了重要贡献。从事应用信息论、语音信号处理和图像处理方面的研究，主持研制“话音数字通信机”。

### 2、能否总结回顾下自己在该领域的研究成果：

1956年，袁保宗教授作为国内首位出国学习信息论的研究生，主要从事声码器7311机和模数模通信研究。1963年，他研究完成了国内声码压缩技术的谐和式声码器，这是现代语音压缩技术在当时最早的成果。1973年，为了填补国内军用通信中的保密及防止窃听方面的空白，他又在国内首次提出了“模数模”话音加密新体制，解决了当时模拟话音无法数字加密的难题，这种体制现已全面应用，并于1978年获全国科学大会重大成果奖。1978年同年成立北京交通大学信息科学研究所，研究语音编码和语音识别与合成，建成国内第一个铁路售票语音应答系统（实验室）。他也率先利用语音识别的成果，开创语音理解的研究，把知识理解运用到语音研究中，完成了汉语语音理解系统。这种研究已经成为以后国际间改善语音识别效果研究的主流方向。1986年国内首次利用原始波形拼接合成汉语语音，由于回避了参数合成语音质量难以提高的技术难点，成为高自然度语音合成的主要趋向。1991年起，又在完成语音识别、理解、合成的基础上，提出了利用知识库进行计算机自动生成语句，实现人机利用语音进行应答，进而于1998年又提出并解决了人机进行自然语音对话时的一些关键问题，为信息处理中的语音应用铺平了道路。

在图像处理方面，袁保宗教授于1983年留学美国，回国后建立了国内第二个图像实验室（采用了法国系

统），研究三维物体识别和计算机视觉定标，其定标精度达到了当时国际先进水平。并用一台计算机控制两台摄像机当做眼睛，将一根绣花针可以准确地穿入80厘米以外的半毫米直径的小针孔中。在此基础上又完成了对三维平面形物体的自动定位及识别、曲面物体的识别以及三维物体的重建等工作。

1988年去英国访问，回国后建立了双眼视觉机器人和运动数据检测控制平台两个主动视觉研究的设备，开展主动视觉研究。

在虚拟现实方面，1990年开始了虚拟现实研究，1994年提出了利用超二次曲面及特征建立的AVR（从真实世界到虚拟现实）方法，实现工件实物直接制成三维计算机图形的核心技术，为计算机视觉与虚拟仿真的结合开辟了新的途径。

在人机交互方面，当“多媒体”一词尚未引入国内时，在完成国家自然科学基金重大项目中，他于1990年率先提出并建立了集声音、图像、图形和文字于一体的“超级智能视听信息处理系统（SIVAIPS）”。这个系统可以完成类似人脑的利用信息进行媒体间的智能转换，并具有视听信息知识融合的新功能，为计算机能听、会看、会学提出了先于国际的理论框架及实现方案，即多媒体视听信息处理结构模型——“圆”模型。这个模型有一个像人一样的大脑，分别控制语音、图像、图形和文字的处理，是一个分布式的网络智能视听信息处理系统。在这个系统平台上，机器能自然地用语音进行人机交互。随着网络技术的发展，他在完成973项目交互式会议电视系统后，又开始了与中科院计算所合作完成国家自然科学基金重点项目“多功能感知机研究”，将SIVAIPS系统，主动视觉以及AVR理论、虚拟现实等研究引入并融合起来，实现了具有智能代理功能及远程虚拟环境的新一代人机自然交互系统（SIVAIPS-2）。2000年之后，随着SIVAIPS-3的研究推进，信息检索、虚拟现实、三维物体重建、人脸识别、应答系统等多方面成果，集成于该系统的演示中，并开放在互联网上，完成智能化自然语音对话信息处理平台。它的功能比现在智能手机的要更多。另外，在1996年863项目支持下完成“交互式会议电视系统”，该系统在863十周年成果展展出，并在郑州铁路局得到应用。

**袁教授：**我根据你写的以上材料，再做了些细节补充。

个人经历的演变发展过程，是从 1956 年开始的。我在 1953 年大学毕业以后就留校任教，当时在国内电信系当老师，然后去听了邮电学院请的一个苏联专家来讲信息论。第一次听，根本不知道什么叫信息论，所以对信息论很感兴趣，听懂了一点。因此，在 1956 年出国时，成为了首批以信息论学习名义出国的副博士研究生。俄罗斯的学位叫副博士研究生，主要是学信息论和一部分语音方面的声码器。声码器是什么意思呢？它叫 Voice Coder 声码器，是把讲话的声音、按口腔的动作进行编码传出去，然后在接收端通过改变模拟口腔的动作跟嗓门，形成声音再发出来，就等于是分析再合成，声音变码，码变声音，就这么一个编解码过程。这种声码器，是当时国际上最高级别的保密机，因为编码方式是未知的，所以声音也就完全加密和难以破密了。做完声码器研究，毕业回国，在国内也做了一个声码器，把国外由电子管做的改进为半导体的声码器，这个成果后来在国内高校成果展览会上进行了展出，这是回国后第一件事。

到了 1973 年的时候，这里要介绍一下 7311 计划。1973 年，尼克松访华之年，北京军区看见人家外国人来，开始的通信是很正常，后来他们把声码器一装上，通信就保密了，你就听不见了。当时，感觉到这家伙挺厉害的，他们用保密机进行通信，我们听不懂他们的通信内容了。我们现在的军用电台，是否也能加一个保密？所以，就找我来合作研究做保密机，叫 7311 机，1973 年 11 月开始做的。当时的电台都是模拟电台，那么模拟电台的声音怎么加密呢？加密是老办法，就是把频段分了好多段，声音分了好多段，然后把它颠倒顺序以后，你不知道顺序的话就保密了。当然这种加密太简单，很容易被破密。因此，我提出了“模-数-模”的模型，把模拟的语音信号变成数字信号，用数字加密后信号再变成模拟信号。由于数字加密很难破密，军用电台的保密性就提高了。要注意 1973 年的时候，数字信号这个概念还没有呢！数字通信还没有这个词，数字化技术也没

有，所以要研究自己做话音量化器。数字量化以后再研究数字加密，用模拟电台传出去，所以模拟的变数字，数字再变模拟传输，到那边以后再模拟变数字，加密后的数字信号，还要再变为模拟信号，在原有的模拟电台上进行通信。“模-数-模”的通信，在当时国内还是第一个，是第一种用数字信号来加密语音信号的军用电台。北京军区的领导都很重视。把成果做成军用电台，在广州进行试验生产。

1978 年的时候，全国科技大会的时候经铁道部批准成立了北京交通大学信息科学研究所，这是国内第一个以信息名义在高校里成立的专职研究所，当时大学里，很少有研究所这个概念。研究所的工作是进行语音编码、语音识别、语音合成的研究，完成第一台国内的铁路售票的语音应答系统，即你问买车票可以打电话来问，我要想买什么车票，它就应答应你，语音应答。由于那时候语音识别技术、语音合成技术都在起步阶段，这就是第一个实验室语音应答系统。

第二个阶段是在 1983 年开始。我去美国留学，到美国进修了不到一年的时间。从语音扩展到图像研究。1983 年回国后，着手建立了国内第二个图像实验室。当时是清华比我们早，图像计算机在这个时候还很神秘，价钱也很贵。我回来以后，搞图像也得搞图像实验室，需要很多钱，而且不单是钱，需要的是外汇，外汇批准要订货，手续还是非常繁琐的。总算在国内清华之后，我们建立了第二个图像实验室。这个图像实验室以法国机器为主，有了这个之后就开始了进行三维物体识别。最开始做的是方块、斜块等各种各样的积木。相机拍下来以后，计算机需要识别这是什么，并进行三维定标，然后开始计算机定标。计算机定标这个事情又简单又复杂，简单的就是你计算两个公式就可以了。复杂的是你要实现很高的精度，那就很难。当时我们这里头做的算比较好的，是第一个博士研究生做的这块工作，基本上定标的精度达到了国际先进水平。当时国际上论文里发表的水平，我们基本都达到了。在一米左右的距离之内，摄像头拍下来，然后确定这一点的三维坐标是多少，误差

在大概是2-3个微米之内。不是毫米级的,是微米级的。有了这样的精度,我们就做了一个演示系统,计算机穿针。直观来说,是打两个双眼。首先,有台机器,前面摆一个牌子。这个牌子上面有一块黑色区域,里面有一个插眼,计算机可以启动针,当这根针打到这一插眼上就停在那。实现了计算机自动穿针的表演。穿针的精度,达到了当时的国际先进水平。

之后,实验室开始搞计算机主动视觉。主动视觉是由计算机控制的机器人能主动地找处理目标,相对于被动视觉有更大的实用性。搞主动视觉的思路也是受国外启发来的。我在英国学习以后,看到人家做的主动视觉机器人头,回来以后我们自己也做了一个改进了的主动视觉机器人头,这个机器人头有一副13个自由度的双摄像头眼睛,可以自由地找到自己要看的目标,进行视觉信息处理。13个自由度就是两个眼睛在XYZ三个方向上的运动和转角,加双摄像头的两个焦距,焦距是两个眼睛分别的焦距,13个自由度,做的比当时国外的机器人头做得好。后来还建立了一个双眼视觉运动测试平台。后来,我在英国问同行剑桥大学教授,他发表过比较权威的论文,我请问他:你通过双眼视觉运动测试,计算得到的结果,包括运动方向与距离的数据,如何去判断它们是正确的?有没有做验证的办法?他没有回答。我告诉他,我们的双眼视觉运动测试平台是实验时,平台先设定好转多少度,什么速度,转到哪里?设定的数据都是已知的,然后进行测试计算,计算的结果可以与设定数据进行对比,来判断计算结果的正确度。我们利用有个运动平台真正做到了主动视觉的正确性。这是我们第二个阶段的图像处理研究。

到第三个阶段就是开始结合搞虚拟现实了。1990年前后,虚拟现实的名词刚刚在中国出现。那时候,马颂德研究员就注意到做这块研究,他是从法国学习回来的,我们非常熟,所以也研究虚拟现实了。当时这个名字都还不知道怎么定义?开始时,搞虚拟现实主要是搞三维立体重建,将左右眼看到的三维立体世界图像,分别显示在屏幕上,通过配戴三维立体眼镜,就能看到三

维立体世界图像了。那个时候我们与医院合作,做了三维的心脏合成、三维的胰岛合成、三维的人头合成,后来发展到三维的静止物体,例如三维观音菩萨、三维阿弥陀佛菩萨的模型等这都做了。三维计算机视觉,看起来简单,做起来很麻烦!现实中的三维物体,用多少幅图像拍下来,可以正确做到三维合成模型?有了物体的三维合成模型,怎样把原物体表面上的纹理张贴到模型的表面?都是三维计算机视觉的关键技术,在当时没有工具软件时,都需要自己研究开发,工作量与难度是很大的!是要贴纹理,要想贴的很好,这还挺复杂的,所以不做实验不知道,做的时候就难度很大了。这是第三个阶段,做了很长时间,做到2000年后还在继续,把虚拟现实成果集成到后来的人机交互系统。

第四个阶段就开始人机交互研究了,计算机听觉视觉从1990年开始发展越来越快了!我1989年到英国去,主要是准备和计划下一步的研究目标!当时,清华自动化系的常迥院士,在做国家基金委的重大项目,这个重大项目是搞语言图像的研究,他说袁老师你再搞重大项目,搞个什么好呢?我在英国学习,我来抓构思,有了构思就有了以后的研究题目:“超级智能视听信息系统”。这里“视听信息”扩大了语言图像的研究内容,“智能”视频信息处理加入了智能的处理,“超级”又不是一般的智能,是一个超级的智能。这个系统就从国家自然科学基金的重大项目立项了,当时我们、清华大学还有上海交大三校合作,我们牵头。清华大学吴佑寿,丁晓青教授,上海交大施鹏飞教授参加。这个项目从1991年开始正式。1990年的时候计算水平还很低,一台计算机搞图像,另一台计算机搞语音,这样有两台计算机合作起来,完成又要听又要看,又要有网络,所以我叫它多机系统,第一个,1993年就完成了,它名为“SIVAIPS-1”。

1997年的时候,进行第二个项目。第二个项目是和中科院计算所高文教授合作叫多功能感知机,我们还继续做我的工作,所以第二个系统做出来了,这个系统我们叫SIVAIPS-2,“2”的意思是原来是分布式的机

器，现在变成集中在一台机器上了，一台机器又可以处理语音，也可以处理图像，网络功能也加强放到里头去了，做完以后，就是我们的 SIVAIPS-2，第二代以后还继续让学生在学，到 08 年的时候，北京市教委立项，批钱继续为学校的一个重点项目。

这样，2008 年以后，在学校的重点项目支持下完成 SIVAIPS-3 是一个网络服务器系统。这由多机变成单机，单机变成了网络，成为第三代了。为什么我要重点讲这些经历？因为它跟现在的 smart phone 功能完全一致了。信息检索、虚拟现实、三维物体识别、人脸识别、应答对话系统等多方面的成果都集成到一起。该系统的演示，集中把它放在互联网网络上，移动用户可以智能化的自然语言对话，和进行各种以上功能的操作。这个信息处理平台，相当于现代智能手机的服务模板。与现在手机上的那些智能相比，我们做得比较早。从 1993 年系统建立到了 2013 年时候，工作比较成熟了。有些功能在智能手机上现在还没实现到。

从以上工作来说，要说贡献的话，第一个是多模态人机交互的“圆模型”。这个模型，我开始在立项会上只是一张“图”，会上经过讨论后，清华大学吴佑寿教授说，这个“图”就叫“袁模型”吧！“圆模型”是我后来改的，把“袁”改为“圆”，不用袁保宗的袁，而用圆周的圆。因为这个模型本来就是一个同心圆的模型。把一个圆分成 4 块，声、像、图、文，四个声、像、图、文板块，代表四种信息媒体。然后把圆分层，分为四层和一个圆心，这样形成四种媒体和四层信息处理的媒体数据。第一层是原始媒体数据层，原始采集到的数据进入第一层，经过信息处理获得信息特征后，进入第二层特征层，特征层的数据再经过模式识别后进入识别层，识别层的数据是媒体的内容了，最后进入融合层，把四种媒体的信息融合起来，形成了理解模式内容的概念。经过 4 个信息处理层次做下来，就是整个信息处理，最后进入到一个“圆心”，相当是大脑。这是一个核心，声像图文的信息在这里进行信息的融合，形成知识，知识存储起来，就可以提供各种应用。以上的“圆模型”就是我们



图 2 袁保宗教授主持开发的 SIVAIPS-3 系统演示界面 SIVAIPS 的构成。

全部信息处理就按照这个模型来进行处理。在这个模型中，我们还提出来所谓“信息模态的相互转换”，它可以把声音信息转变成图像信息，图像信息转变成文字信息，文字信息转变成声音信息，这种互相转换，对我们理解世界和重建很有用！以上模型是一个多模态信息处理系统，可以实现信息理解，转换，融合及应用。当然，在完成这个系统的过程中，我们合作老师阮秋琦教授，唐晓芳老师和二十来名博硕士研究生在语音图像、三维重建、信息检索、信息融合方面做出了大量的成果！

1996 年时候我们还做了一件事情，就叫“交互式会议电视系统”。这是 863 项目，主要是解决当时国内会议电视多是进口的，我们要建立国产的交互式会议电视系统，这项研究是我所裴正定教授领导完成的。后来在 863 的 10 年成果展上展出，展出以后生产的设备被郑州铁路局使用，解决了当时的应用。

**明：**通过这个领域的演变过程能否为当下青年学者提供一些研究方向演变的经验和规律？

**袁教授：**说到有什么经验，规律？第一个就是**选题**！做科研计划选题很重要，要符合国家的需求，这是根本一条。没有国家的需求，光是自己凭兴趣那不行。我们刚才说的这些项目都是国家的重点项目，这点很重要的。现在作为教授来讲，你必须有国内的重大项目，你才能

做，没重大项目你怎么能有重大成果？这是不可能的，所以要争取这些国家级项目。

第二就是要有对科技前沿的**好奇心**。反正，我这个人比较怪，对新的东西很感兴趣，想再听一听是怎么回事？哪怕不是自己专业、旁边专业的也要去听听，听听之后也会说：哎，这个好。好奇心很重要，没有好奇心了，以后你就没进取心了。所以第二点一定要有前沿方向的爱好。这种好奇心是我们研究的动力之一，是内在的动力。

第三就是个人要有**愿意实践，想自己来做一下**！算法与程序尽量都要自己做一下，一般学生做好了，他给我汇报，汇报完了以后，我也得拿他的程序运行一遍，实践一下后脑子里印象深一点。现在，看学术论文的话，往往程序都带着的！都在网上放着，是吗？下载后，我自己再走一遍很容易。譬如，近几年热门的深度神经网络现在用的 CNN (Convolutional Neural Network) 和受限玻尔兹曼机 RBM，这些我都做过一点。自己实践一下之后，你心里就踏实了！

**明：**能否再谈谈您关于学位点的贡献？

### 学位点建设

袁保宗教授既是一名信息科学的专家，同时也是一位资深教授。大学毕业以后的几十年中，他一直辛勤耕耘在教育战线上，可谓桃李满天下。听过他讲课的学生都有一个共同的体会，那就是不管多么深奥的理论，多么复杂的难题，经过他一讲，都是条理清楚，明明白白，深入浅出，既严谨，又幽默，是一种享受。因为在他的教师生涯中有一个信念，那就是“要把我的知识毫无保留地教给我的学生，让他们听懂、掌握。”为了这个信念，在课下他要花比别人多几倍的时间备课，甚至连讲课的表情、手势、板书的位置等都要提前设计好。在他的培养下，成长为包括硕士、博士、博士后在内的高层次人才已有百余名。他于1994年为北方交通大学接收了第一位外籍（法国）博士后研究人员，1995年为北方交通大学培养了第一名外籍博士生。他的学生分布在世界五大洲，有的已经成为了教授、博士生导师、学科带头人、企业的

技术领导等，成为了信息科学领域中强有力的生力军。为了能更多更好地培养高科技人才，袁保宗教授积极建设学校的博士点和硕士点，在他的努力和共同参与下，先后创建了博士学位二级学科授予点3个，一级学科授予点1个，博士后流动站1个。1986年，“铁道运输自动化与通信”学科点被批准为国家级重点学科，1990年在原有博士点的基础上，申请批准了“通信与信息工程”博士点，1996年又再次申请获准了“信号与信息处理”博士点，1998年，国家学位委员会批准通过了“通信与信息工程”一级学科授予权，1999年批准设立了“通信与信息”的博士后流动站。在他的带领下，信息学科的学位点建设发展迅速，以信息科学研究所为基础的博士生导师队伍迅速成长，他积极吸收来自数学、物理以及计算机系的教授，聘为博士生导师。导师队伍的成长带动了博士研究生队伍的扩大，相应的研究实验室于1999年被铁道部批准为部级开放重点实验室，为培养铁路信息技术的高级科技人才创造了良好的环境。

**袁教授：**关于学位点建设在你整理的材料里头都有了，我就不说了。

**明：**能否再谈谈您在学术交流方面的成就和经验？

### 学术交流

为了发展中国数字信号处理领域的新技术，新成果，1979年，在他与学部委员罗沛霖、程民德和常迥教授的共同倡议下，参与创建了中国电子学会信号处理分会，袁保宗教授长期担任学会的理事长及副理事长，他是第五届委员会的理事长。在他的组织下，开发出版和推广了两个信号处理通用程序库，为国内研究单位起步信号处理工作起到了推动作用。20多年来，在他领导下，广大同行专家通过召开每年的学术会议，通过在学会主办的学报《信号处理》上发表论文，互相交流学术思想，进行学术讨论，推动了全国信号处理科技的发展和应用。1985年在北京成立IEEE北京分部，1992年英国的IEE，他推动了将这两个科技协会都引入中国，现在IEEE的北京分部就设立在北京交通大学信息科学研究所。他先后担任了IEEE北京分会信号处理分会和计算机分会的主席。另外，自20世纪80年代初，他还担任了《电子学报》、《通信学报》等学报的常务编委，国际《人工智能工

程应用》、国际《电工教育》等学报的编委，以学术影响来推动中国信号处理技术走向世界前列。

**袁教授：**我根据学术交流里头的材料，增加了一个数字信号处理的技术。

数字信号处理这个词，本来当时还没有，后来奥本海姆出了信号处理这本书，就有了数字信号处理的名词，这本书我们后来也看了、翻译了，这个书就是我们信号处理的开始。在1979年的时候，在罗沛霖、程民德、常迥三位教授的倡导下，参与建设了中国电子学会的信号处理分会。开始是常迥当理事长，后来常迥当了一届以后，他说袁老师你当吧，我就继续担任这个分会的理事长。

在我们学会组织下，出版了一个当时来讲应该是一个通用的程序册，讲的是信号处理是怎么回事，怎么做，计算机怎么来完成计算？都有些什么算法呢？当时根据一本外国的书的基础上，把中国高校组织起来，也出了一本国内的程序册，叫信号处理通用程序库。有了这个库以后，如果你不会做信号处理，没关系，你用库里的程序吧！所以，这本书里的库，当时在国内推广信号处理起步方面起到很大的作用。后来，为了能让中国的信号处理研究跟上国际的步伐，为了让更多的国内学者与国际同行进行学术交流，自1990年起，又组织由中国牵头召开的国际信号处理会议（ICSP），并担任会议主席和程序委员会主席，每两年一次，至今已举办了十五届。今年是ICSP会议召开的30周年，ICSP2020由于新冠疫情的原因，会议改为互联网线上线下结合召开，向全世界直播，海内外学者在线参加会议。开幕式上，大会还向国际上30年来积极支持大会的研究人员、学术机构、领导部门、老师及工作人员都发了奖，表示鼓励，感谢对我们工作的支持。该会议在国际上取得了成功，产生了重要影响，已经被公认为与美洲、欧洲并列的，在亚洲召开的信号处理国际系列会议之一，使中国信号处理领域的学者不出国门就可以与世界各地的同行专家互相交流学术思想。

再补充这么一点，1997年底应日本信号处理的专家邀请，组织了日中专家信号处理会议，组织了国内的专家赴日讲学，并担任了大会副主席，我们中国去了10个人，10个方面各方代表到日本去做一些交流，交流了两国的技术成就，扩大了中国信号处理学界的国际影响，日本方面也给我们中方做了交流及参观。

**明：**您能否再谈谈学术前沿如何把握呢？

### 学科前沿方向

长期教学科研的实践，使袁保宗教授悟出了一个信条：“科学问题上不得半点虚假”、“知识愈浅，自吾愈大；知识愈深，自吾愈小”、“初学三年，天下去得，再学三年，寸步难行”。随着科学技术的发展，袁保宗教授深知，只有不断结合中国实践，吸收国际先进科技，才能不断充实自己，培养好学生。他在1983年去美国进修期间，详细研究了国外大学博士论文的格式、内容、要求和答辩等问题，并用到了自己培养的学生要求中。1985年始，他自己订阅了十几种国外IEEE的学术刊物，从中吸取最新研究成果，从刊物论文中吸取新思想，丰富新知识，帮助提出新问题，看清学科发展新方向。他还积极从参与的社会学术活动中吸取新动力。他参与国内5个国家级重点实验室的建设工作，担任中国科学院声学研究所“国家声场声信息实验室”副主任、北京大学“视听信息处理国家实验室”学术委员会副主任、中国科学院自动化研究所“国家模式识别实验室”学术委员会委员、清华大学“国家智能技术和系统实验室”委员、原武汉测绘大学“国家测绘遥感信息技术实验室”学术委员会委员。从参与这些实验室建设过程中，他一方面为他们把握学科方向，另一方面亦把他们先进的经验带回自己的学校，教会学生向先进单位学习。正是由于他勤于实践的作风，使他与学生讨论理论问题时能更加深入浅出，研讨实践方法时更加切实可行，与国内外学者进行学术交流时提出问题能更切中要害。谦虚谨慎，勤奋好学是他知识常新的源泉。

**袁教授：**我再说说学科前沿方向，你提到的，**长期教学实践**，使得袁保宗教授悟出了一个信条，科学问题上不得半点虚假，知识越浅自我越大，知识越深自我越小，这是你写的。我这里再加一点，叫**初学三年天下具得**，

**再学三年寸步难行**，这个是我们小时候说的，后来也确实是这样的。我们工作以后，初学三年天下具得，就像研究生这个阶段。我研究生毕业后，读了点信息论，读了一点优化理论，一来一算，感觉心里好像很踏实。一出来就说这不就是最优化嘛，一算就是了，看着都很简单，叫初学三年天下去得，你刚刚起步。再说再学三年寸步难行，你再过三年以后，这个问题不是你想的那么简单，复杂的很，你寸步难行了，所以说说话要虚心点吧，不要随便说话，你们当教授的更要注意。

这里再补充一点，我在 1983 年去美国进修的时候，详细的研究了国外大学博士论文的格式、内容、要求、答辩。这个可能国内也比较少。因为 1983 年的时候，博士生还没开始呢。那时候学校里头定博士生导师了，这博士生是什么样的？不知道那怎么办？我们以前在苏联做过论文，这个论文叫副博士论文，不是博士论文？那么当时去美国后，就在那读了 20, 30 多本博士论文，美国的博士论文，这一篇一篇的看，你看了他的博士论文以后，把他的格式、内容、要求、答辩这些学会，这些学了以后回来写了一个提纲要求，然后把这个东西学回来作为自己研究时的版本。所以有一个参考价值，这个应该如何也知道了。

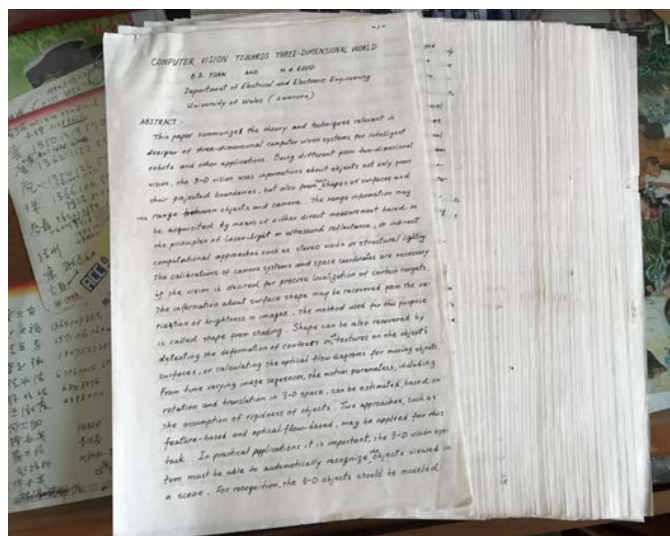


图 3 袁保宗教授在美国访学时手抄的美国博士论文

**明：**有什么值得回忆的关于该领域有趣的轶事？

**袁教授：**关于本领域有哪些趣事。我觉得我可以给你补

充一点有趣的轶事，或者不叫它有趣的，是有点意思的。我们在学习中常常有些事情很难懂，书上一句话很简单过来了，可是你要懂得它，还是要动动脑子的！我有这么一个习惯，经常从生活实践当中寻找解释科学原理的“实例”，和从生活实践当中去找解决科学问题“方法和原理”。

这里举两个例子，第一个例子是：我们在读文章里常常这么说，“低维空间里头的非线性问题，经过变换到高维空间就可以用线性问题解决”，是不是？这个谁证明过？谁解释过，我觉得没一个人解释过。我这怎么办？我就说很简单地这样去理解：这里有一个容器，容器里有条手绢，手绢上下都是沙，上面放点啥，团起来，手绢是弯弯曲曲不平的，可以看成非线性的。你把手绢四个角一撑，上面上去了，下面下来了，撑平了就成线性了，平面是线性的。这样跟学生讨论问题，大家觉得很容易懂了。

再一个是关于黎曼流形 Manifold! 它的测地距离与欧氏距离之间的关系！它与对称正定矩阵的关系！由于涉及微分几何内容很不容易理解。我把对称正定矩阵的流形上的测地距离与它切平面上的欧氏距离看成如海底与海面上点与点之间的距离，这样就容易理解了！

**明：**您如何看待这一领域的发展：过去，今天与未来

**袁教授：**对我们搞研究来说，最重要的还是**创新**。创新必须先学习，获得思维的启发。你不学习你是没法创新的。所谓学习，就是先学人家的东西，取长补短，你才能去做创新。另外，从生活实践当中也可以去找类比，来解决你的难题。生活当中就有类比，类比可以帮助我们找到创新的办法。

**学会从批判中来学习也是有帮助的。**我们以前当学生的時候，总希望老师有什么指导？我在苏联学习时，当学生的時候，两个礼拜要向导师汇报一次学习情况，老师在前面坐着，我在黑板前面讲，导师在前面听，听完了以后，他就要问你：这篇文章正确的地方在哪里？特点在哪？错误的地方是哪里？要你找几个毛病出来。

我们找的一些毛病问题往往存在于文章中的“假设”上！看看文章的立据，计算的一些条件，结论是否全面等等？通过这种批判式学习可以提高我们分析和解决问题的能力。

最后，还有一条很重要，就是要学会“创新思维”！什么叫创新思维？这是我在学习“信息论”时学来的！香农信息论是很难懂的。我在留苏回国以后，第一个就是讲信息论课，最难懂的就是香农信息论的定义，它的证明。香农发展了信息论，而信息论的诞生就是新思维逻辑的创新。现在还没一个什么任务，像信息论这样一个诞生过程的描述。第一，确定信息的基本概念，信息论以前没有信息这个概念，只有概率论没有信息论。他把概率论提高到信息论。香农把“随机事件”的出现，可以给人们提供一定的“信息”，随机事件与信息联系起来。同时还给出这个信息的定量“信息量”和单位。在这基本概念上的基础上又给出了一套公理性的定量评价。所谓公理性，就是没有证明的。公理性必须承认，譬如信息一加信息二，信息三一定要大于信息一和二。将公理性定义为公理，然后推出来计算公式。第三是什么？解释各种概念的量纲和它的极限量，最优可以优到什么程度，最坏可以坏到什么程度？给出了一系列公式。这就是香农信息论给出个极限定理。从此它通过计算，指明不论语音，图像，视频等都可以大大地进行数据压缩，引导了无数研究人员开展研究。我们说一辈子实际都是实现了信息论的香农定理里指出来的这些方向，到现在为止基本都实现了。

**明：**您还有什么需要告诫后辈的，比如硕博士学生：如何阅读文献、撰写论文、提升对科学问题凝练的能力、团队协作能力、演讲和表达能力？再比如青年教师：如何迅速确认个人发展方向？

**袁教授：**给研究生们说，首先，多方面的“知识要整理总结”，零星的要变成整体，因为这是一条学习路径。现在我们学习，我也感觉到学习的知识过分扩散化了。知识特别分散，你不去总结它，那是零星。零星的没什么用，是吧？所以现在数据时代，把所有数据收集，进行

融合整理，能够计算，得出来你要的知识。

第二个，“**计算工具要不断积累**”。你们做博士生的可能知道的工具比我多，但我也是在积累，例如计算工具——Matlab。现在是python，还有其他好多好多算法，平台嘛，我们中国人没有做，所以这个还是要积累，要归类，要使用，要建立自己的工具库给别人使用。现在我们国内有几个学校开始做这个事，国家模式识别实验室有这个计划，现在他那已经有一套做出来了，给大家用，主要是关于视觉听觉。现在清华大学胡事民也在做。我们基本计算的工具，因为没有自主产权，一旦现在不让你用了。要自己准备工具。如果你能够有自己作一套工具，哪怕很简单的东西，大家能够用的话，那你就不得了了。

第三点，“**演讲表达能力要苦练，勤学苦练**”。你们没有这一课，在国外还有叫presentation。老师教你们presentation怎么做吗？这谁教过你应该怎么讲？没教过你们，怎么讲得让大家听得津津有味，对吧？这次我开了三天国际会议，100多个报告我都听了，三天时间在那坐着听的，没一个讲得好好的。怎么叫没一个讲得好的？他只会说题目，然后包括有一个简述，序言，然后公式，然后优化，然后结论这实验结论，就这么几个。就没有讲这个事情，事出有因，为什么要做，怎么做？别人为什么做不出来？没有讲大家愿意听的部分，最枯燥的把它拿出来了，最精华的东西没有。这个结论性的东西拿出来了，这个结论是人家不懂的，一点用没有。你应该给人家讲启发性的东西，思路在哪里？然后你的方法，轻描淡写说一说就行了。方法做完了以后好处在哪里？讲清楚一点，人家听，这是管用的。就是说怎么做presentation，大家自己要好好的学习一下。

**明：**您认为今后哪些方向值得重点关注，比如：基础理论、工程应用、人才培养。

**袁教授：**数学是最基础的理论，一定要反复学习。任何现象都要从数学来做个解释！

我觉得，数学是最重要的基础，一定要反复的学习，

任何的现象都可以归纳到数学来解释。这是我的一个认识。因为现在来看去数学，无非就用了一个优化公式，那极限在哪里，你们不知道对吗？现在有了深度学习网络之后，大家都变得不那么重视数学，全都去改网络结构，没有数学没有办法。优化网络也是这样，神经网络深度学习的过程，它也是需要优化公式。但是，这种认识有片面性，因为人的思维，不仅是以公式计算来的，人的思维是各方面数据推导综合起来，光一个公式不能解决所有问题。神经网络不是万能的，解决不了所有问题。可能要许多神经网络联合去做一件事情，做出的结果再融合得到新研究结果。

数学作为基本要反复学，底层东西都是数学的，比如，微分方程是描述世界运动规律的数学。电磁波理论整个就是麦克斯韦方程！所有电磁问题都可用麦克斯韦方程解释，但是麦克斯韦是个微分方程，微分方程现在课堂上也不讲了！现在，图论也不讲了，就讲图的聚类方法！几何问题的课程也没有了，高等几何，微分几何都没有了吧！逻辑学更没有了。很多数学工具都没用到！实际上，微分方程是最基本的，非线性动力学解决的是优化问题，结果收敛到某一个点上，就是系统的稳定点。我自己数学没有学好，所以，经常从网上下载一些数学

书，慢慢自己学习看看！

下面，叫我说一下有什么看法？对于我们现在的时代，我认为是这样的：我们走过了信息、数据、网络 and 智能时代，我们出生的是信息时代，近几年是数据时代，现在最近的阶段是大数据，物联网，5G 时代。这个时代样样都要网络，一天到晚 AI。所以，可以说我们走过来的叫信息、数据、网络、AI 时代。今后怎么办呢？什么时候可以走到下一段？下个什么时代是什么？我就想，想不出是怎么了！我只能在这感慨一下，什么时候可以对思维、知识进行量化计算了，新的时代可能到来了！思维的定量，知识的定量，计算机能够对思维和知识进行计算的时候，这些是我们下一阶段很难解决的问题，目前来看还没有什么很好的方法。现在讲 Ontology，就是知识的层次。Ontology 可能这是一条考虑的路，就是在 Ontology 的基础上进一步做定量的指标，数学分析。这可能是一种思路，只是我考虑的思路。当然，人工智能的社会伦理与公约也不能不加以关注！！

**注：**文中粉色框内楷体部分主要摘自百度百科袁保宗教授 (<https://baike.baidu.com/item/袁保宗/5646639?fr=aladdin>)

**责任编辑** 明悦 张军平 贾熹滨



## 袁保宗

袁保宗教授是信号与信息处理专家，资深教授。1950 年-1953 年，就读于北方交通大学。1960 年获苏联列宁格勒铁道工程学院技术科学副博士学位。曾任北京交通大学教授及信息科学研究所首任所长，电子工程与信息技术学院院长。担任科技部国家重点基础技术研究项目“973”信息领域咨询专家，国家自然科学基金通信信息领域评审专家，国务院学位委员会第二、三、四届评议组成员，中国电子学会和中国通信学会常务理事，中国电子学会信号处理分会主任委员等，1986 年获全国五一劳动奖章等各种奖项。1992 年英国国际电机工程学会（IEE 会士；北京分会副主席）。2000 年获 IEEE 国际千年奖（注：国内一共近 10 位）和奖章。2018 年 11 月 24 日，获 2018 年度 CCF-CV 终身学术贡献奖。

# COMPUTER VISION NEWSLETTER

01 2021  
总第 27 期



## 计算机视觉专委会简报



CCF 计算机视觉  
专委会