

# 人工智能之计算机视觉

## Research Report of Computer Vision

2020年第8期

清华大学人工智能研究院

北京智源人工智能研究院

清华-中国工程院知识智能联合研究中心

2020年10月

## 前 言

计算机视觉（Computer Vision, CV）作为人工智能（AI）的核心技术之一，在过去的三十年里发展迅猛，应用范围遍及工业、农业、军事、国防等多个领域。与人类相比，机器更具优势，它不需要像人类那样依赖可见光，而是利用传感器就能更清楚地看世界。但从发展角度看，计算机视觉技术本身还在发展，还有许多理论、算法尚需完善，计算机视觉的应用范围也还远没有达到普及的程度，此项技术蕴藏的潜能亟待开发利用。

本期 TR 报告我们选取计算机视觉作为主题，围绕计算机视觉的基本概念、技术发展、人才概况、产业应用和热点趋势五大方面进行深入挖掘。报告的论文、国家自然科学基金、趋势数据均来自于清华大学唐杰教授自主研发的“科技情报大数据挖掘与服务系统平台”（简称 AMiner），利用人工智能、大数据分析、知识图谱、自然语言处理等技术，并结合文献计量学等情报学方法制作生成。

# 目 录

图表目录.....	IV
<b>1. 概述篇</b> .....	3
1.1 计算机视觉的概念.....	3
1.2 计算机视觉经典任务.....	3
1.3 计算机视觉的产生与发展.....	6
1.4 计算机视觉的机遇与挑战.....	6
<b>2. 技术篇</b> .....	9
2.1 图像增强.....	9
2.1.1 基于多尺度分析的图像增强.....	9
2.1.2 数学形态增强.....	9
2.1.3 卷积神经网络增强.....	9
2.2 图像分类.....	10
2.2.1 单标签分类.....	10
2.2.2 多标签分类.....	10
2.3 图像检测与定位.....	11
2.3.1 物体定位.....	11
2.3.2 关键点检测.....	11
2.4 图像分割.....	11
2.4.1 语义分割.....	11
2.4.2 实例分割.....	12
2.4.3 全景分割.....	13

---

2.5 目标识别.....	13
2.5.1 3D 目标识别.....	13
2.5.2 点云目标识别.....	14
2.6 专利申请情况.....	15
2.6.1 全球专利申请概况.....	15
2.6.2 中国专利申请概况.....	16
2.7 国家自然科学基金支持情况.....	17
<b>3. 人才篇</b> .....	19
3.1 计算机视觉学者概览.....	19
3.1.1 全球学者概况.....	19
3.1.2 中国学者概况.....	22
3.2 计算机视觉代表性学者介绍.....	23
3.3 计算机视觉代表性团队介绍.....	50
<b>4. 应用篇</b> .....	57
4.1 城市公共安全.....	59
4.2 政务民生.....	60
4.3 金融服务.....	60
4.4 新零售.....	61
4.5 产业应用的未来.....	62
<b>5. 趋势篇</b> .....	63
5.1 技术趋势.....	63
5.2 国家趋势.....	63
5.3 机构趋势.....	64

5.4 技术发展面临的挑战.....	65
参考文献.....	67

AMiner

---

## 图表目录

图 1 识别任务.....	4
图 2 运动分析.....	5
图 3 场景重建.....	5
图 4 语义分割示例.....	12
图 5 实例分割示例.....	12
图 6 基于模型的 3D 目标识别方法流程.....	14
图 7 计算机视觉领域专利申请情况.....	15
图 8 全球计算机视觉专利 TOP 3 国家年变化趋势.....	16
图 9 中国计算机视觉领域专利申请量 TOP 10 省市.....	16
图 10 中国计算机视觉专利 TOP 3 省市年变化趋势.....	17
图 11 国家自然科学基金分布情况.....	17
图 12 国家自然科学基金项目量前十的依托单位.....	18
图 13 计算机视觉全球顶尖学者分布.....	20
图 14 计算机视觉学者数量 Top 10 国家.....	20
图 15 计算机视觉学者 h-index 分布.....	21
图 16 计算机视觉全球学者迁徙图.....	21
图 17 计算机视觉学术机构对比.....	22
图 18 中国计算机视觉领域学者分布.....	23
图 19 计算机视觉领域技术趋势图.....	63
图 20 计算机视觉领域国家趋势图.....	64
图 21 计算机视觉领域机构趋势图.....	65
表 1 中外国家合作统计.....	23

---

## 报告说明

### 1.数据来源

本报告中与计算机视觉领域相关的人才数据均来自于 AMiner 系统。系统支持研究者信息抽取、研究者社会网络关系识别、研究者能力图谱、审稿人智能推荐等功能,提供研究者和研究领域的全面知识,为科研管理和服务提供有力支撑。平台自 2006 年上线以来,经过十多年的建设发展,已建立运作良好的数据采集及集成更新机制,收录论文超 3 亿篇,专利 1 亿项,学者 1.3 亿位,其中超过 50 万的学者被人工标注与审核,吸引了全球 220 个国家/地区 1000 多万独立 IP 的访问,年度访问量 1,800 余万次。

### 2.学者及研究领域筛选方法

本次报告中的人才和技术篇采用大数据挖掘技术,对计算机视觉领域内的学者信息进行深入挖掘,参考 h-index、发表论文数、论文被引频次等指标,对学者信息进行筛选,比较和分析了计算机视觉领域人才在全球和国内的分布概况,领域的技术研究发展趋势,以及技术领先国家趋势等。

(1) 由计算机视觉顾问组推荐期刊/会议列表和领域关键词,①期刊/会议: IEEE 国际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR)、IEEE 国际计算机视觉大会 (IEEE International Conference on Computer Vision, ICCV)、欧洲计算机视觉国际会议 (European Conference on Computer Vision, ECCV)、IEEE 模式分析和机器智能学报 (IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI)、国际计算机视觉杂志 (International Journal on Computer Vision, IJCV)、IEEE 图像处理汇刊 (IEEE Transactions on Image Processing, TIP)、计算机视觉和图像理解 (Computer Vision and Image Understanding, CVIU)、模式识别 (Pattern Recognition, PR)、模式识别快报 (Pattern Recognition Letters, PRL); ②领域关键词: 计算机视觉 (Computer vision)、图像和视频采集 (Image and video acquisition)、图像识别 (Image recognition)、模式识别 (Pattern recognition)、图像理解 (Image understanding)、图像分类 (Image classification)、目标检测 (Object detection)、目标跟踪 (Object tracking)、语义分割 (Semantic segmentation)、实例分割 (Instance segmentation)、卷积神经网络 (Convolutional neural networks)、三维视觉 (3D

---

vision)、光流估计(Optical flow estimation)、景深估计(Scene depth estimation)、人脸识别(Face recognition)、图像生成(Image generation)。

(2) 通过 AMiner 大数据平台对近 20 年(2000~2019 年)发表在推荐期刊/会议的论文进行采集和清洗,并对论文作者信息进行深度挖掘,从中选出了与计算机视觉领域关键词相关的 50,073 位学者,再按照学者的 h-index 进行排序;

(3) 综合运用知识图谱、自然语言处理、可视化、文献计量学等技术手段,基于论文和学者数据,分析得出计算机视觉领域的技术研究发展趋势,以及技术领先的国家、机构趋势。

### 3.代表性学者画像

“学者画像”是 AMiner 平台的核心服务功能之一,其特色在于除了提供专家学者如姓名、单位、地址、联系方式、个人简介、教育经历等个人基本信息之外,还利用团队多年的命名排歧相关技术,建立了较为完善的学者-论文映射关系,提供学者学术评价、研究兴趣发展趋势分析、学者合作者关系网络等分析挖掘信息。

# 1. 概述篇





## 1.1 计算机视觉的概念

计算机视觉（Computer Vision, CV）是指用计算机实现人的视觉功能——对客观世界的三维场景的感知、识别和理解<sup>[1]</sup>。它是一种典型的交叉学科研究领域，包含了生物、心理、物理、工程、数学、计算机科学等领域，存在与其他许多学科或研究方向之间相互渗透、相互支撑的关系。

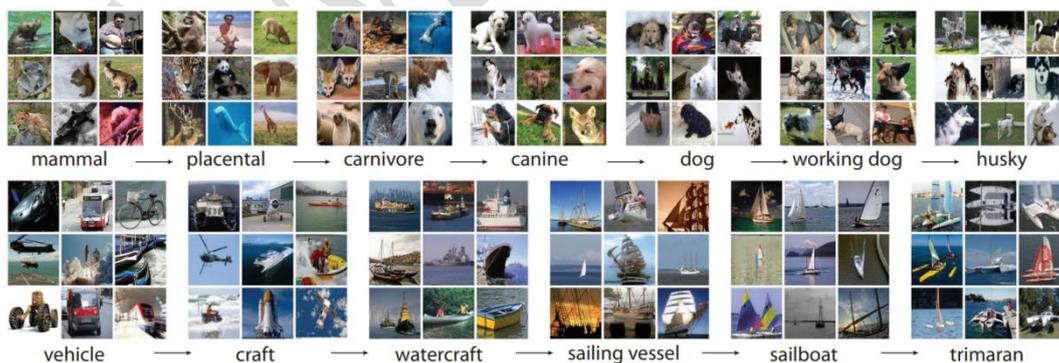
## 1.2 计算机视觉经典任务

近几十年来，随着计算机硬件、机器学习以及模式识别技术的快速发展，特别是伴随着深度学习技术的崛起，计算机视觉的研究得到了飞速发展。目前，计算机视觉已经在光学字符识别、人脸检测与识别、物体检测等领域的多个大数据集评测中接近或者已经超过了人眼的性能。概括来说，计算机视觉主要有以下几个经典任务。

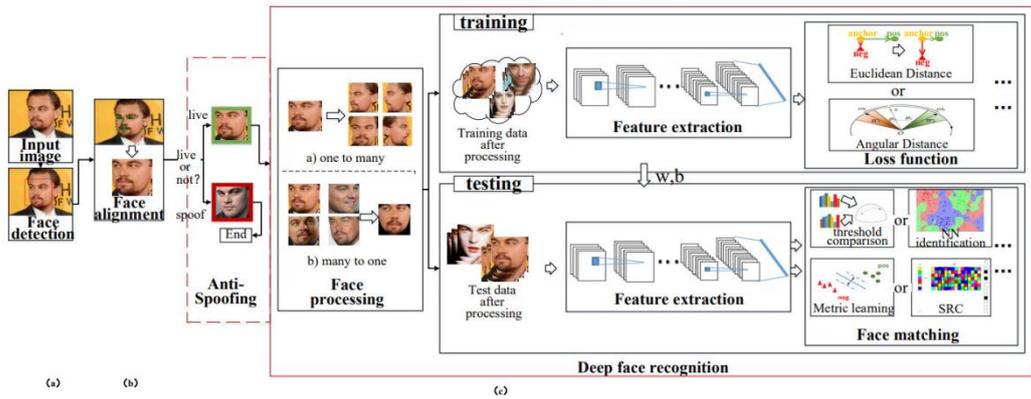
- 识别任务

计算机视觉的经典问题是确定图像是否包含特定的物体、特征或活动。识别任务又可以细分为图像分类、标识和目标检测。图像分类指的是将图片归入事先设定的类别中，例如识别图片中的物种识别<sup>[2]</sup>；标识指的是获得人或物体的身份信息，例如人脸识别<sup>[3-4]</sup>；目标检测指的是检测图像中存在的感兴趣的一类事物，例如获得图中车辆、行人的位置信息<sup>[5]</sup>。

(a) 物种识别任务<sup>[2]</sup>



(b) 人脸识别任务<sup>[4]</sup>



(c) 目标检测任务<sup>[5]</sup>

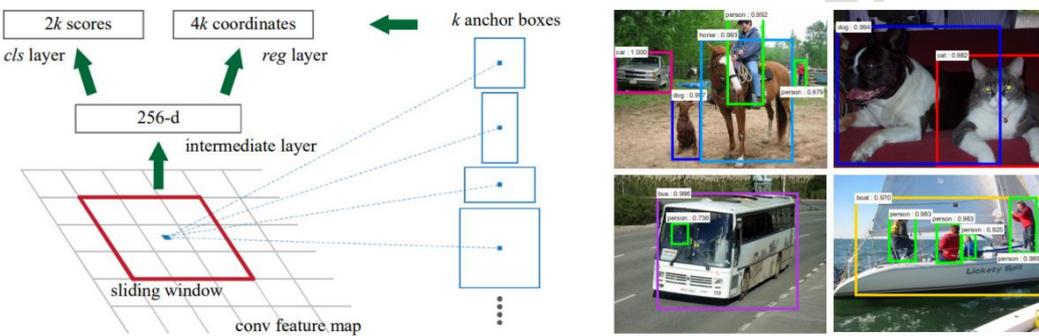


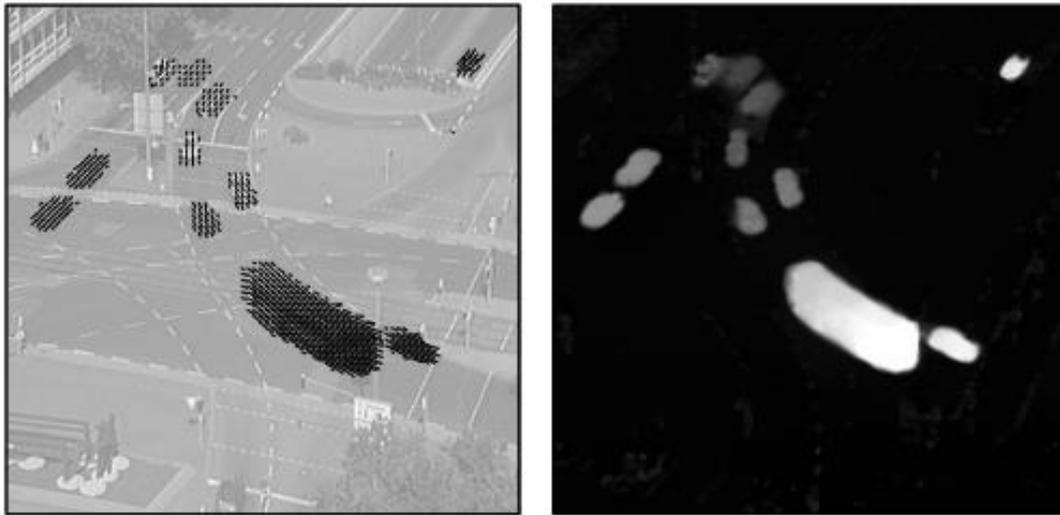
图 1 识别任务

● 运动分析

这类任务的目的是估计图片序列中物体的运动状态。例如，根据拍摄视频估计相机的运动<sup>[6]</sup>，追踪监控视频中的物体<sup>[7]</sup>，计算光流——即图像上点在下一张图像的位置<sup>[8]</sup>等等。

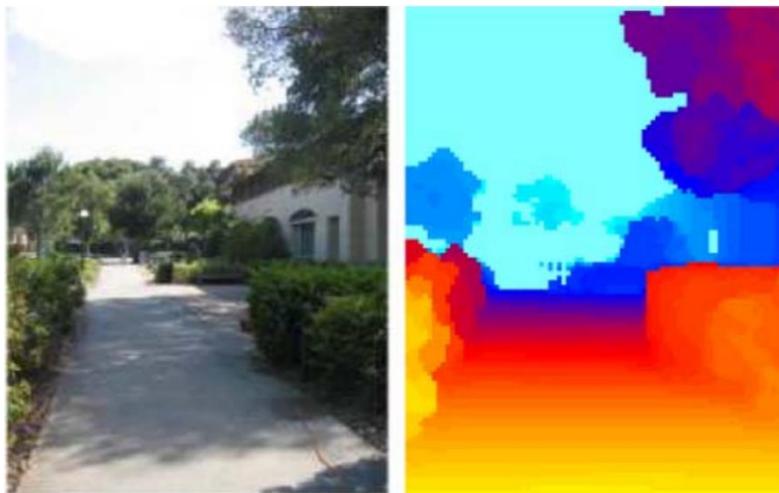
(a) 追踪监控视频<sup>[7]</sup>



(b) 计算光流<sup>[8]</sup>图 2 运动分析<sup>[7]</sup>

- 场景重建

基于已有的一张或多张图片建立场景的三维模型<sup>[9]</sup>。一种简单的情况是确定场景表面的一组点，更复杂一点可以是多面体模型。场景重建可以在不借助运动和扫描的情况下构建三维模型，这在虚拟现实场景下可以获得广泛应用<sup>[10]</sup>。

图 3 场景重建<sup>[10]</sup>

- 图像恢复

图像恢复<sup>[11]</sup>的目的是去噪。相机成像是通过光线打在光学传感器上实现的。传感器和相对运动都有可能带来噪声。计算机视觉的去噪的方法包括滤波器<sup>[12]</sup>和深度学习模型<sup>[13]</sup>。

## 1.3 计算机视觉的产生与发展

纵观计算机视觉发展历史，其发展与很多学科息息相关<sup>[13-14]</sup>，大致可分为以下四个阶段。

### 第一阶段 马尔计算视觉

1982年，大卫马尔（David Marr）的《视觉》一书在计算机视觉领域中起到了关键性的作用，它标志着计算机视觉正式成为一门独立的学科。马尔的计算视觉分为三个层次：计算理论、表达和算法以及算法实现。马尔认为，大脑的神经计算和计算机的数值计算没有本质区别，而从现在神经科学的进展看，“神经计算”与数值计算在有些情况下还是会产生本质区别。

### 第二阶段 主动视觉与目的视觉

学术界几位教授对马尔视觉计算理论提出了反对意见，认为缺乏主动性、目的性和应用性。但由于这段时期没有过多进展，对后续计算机视觉的发展影响不大，因此很多时候没有把这一阶段单独列出介绍。

### 第三阶段 多视几何和分层三维重建

其中代表人物包括法国的 O. Faugeras、澳大利亚国立大学的 R. Hartley 和英国牛津大学的 A. Zisserman，研究重点是如何快速、鲁棒地重建大场景。

### 第四阶段 基于学习的视觉

最后来到了当代计算机视觉的阶段，基于学习的视觉。在此阶段中，文献大体上分为两个阶段：一是以流形学习为代表的子空间法，二是目前以深度神经网络和深度学习为代表的视觉方法。

## 1.4 计算机视觉的机遇与挑战

自 20 世纪 60 年代开始，计算机视觉取得了长足的进步，特别是在图像分类、人脸识别、目标检测、医疗读图等任务上逼近甚至超越了普通人类的视觉能力。计算机视觉的所面临的机遇与挑战主要表现在。

（1）计算机视觉迎来了前所未有的关注和接踵而至的投资热潮，这些关注既来自风险投资公司、互联网公司，也来自各级政府，据艾媒咨询数据显示，2018 年中国计算机视觉市场规模为 155 亿元，较 2017 年增多了 87 亿元。2019 年中国计算机视觉市场规模达到 450 亿元，2020 年达 780 亿元，2021 年将突破 1000 亿元，达 1120 亿元（<https://data.iimedia.cn/>）。另外，在中国人工智能创业

公司所属领域分布中，计算机视觉领域拥有最多创业公司，包括商汤、旷世、云从、依图等众多硬科技公司。其中，商汤、旷世、云从更是登上了美国的“实体清单”，作为推动中国高科技发展的中坚力量，要想在国际舞台上拥有更多的话语权和主动权，就必须手握核心技术，拥有自主产权的高端产品。

(2) 计算机视觉是一类相对技术发展较成熟、应用场景多样、误判容忍度有不同要求的可深度发展的领域，很难出现互联网行业那样赢者通吃的局面。在应用层面上，移动互联网/安防领跑，零售/物流跟进，医疗/无人驾驶有待成熟，落地的速度开始出现分化。社交、咨询、游戏、电商等移动互联网场景，以及门禁等安防领域，因为数据比较容易获得，以及对误判容忍度相对较高，发展非常迅速。预计新零售、物流、制造业等企业场景以及家庭安防等家居场景开始成熟。另一方面，无人驾驶及医疗对辨别的准确性要求高，数据复杂程度高，短期很难实现大规模商用。

(3) 尽管美国在计算机视觉领域一直处于前沿，但中国已悄悄开始“超车”，大批华人科学家，如李飞飞、汤晓鸥、孙剑、何恺明等已在学术和产业界居于领跑者地位，华人计算机视觉领域精英们正在撼动着美国在人工智能领域的领导地位，并逐渐获得了世界的肯定，使中国迎来了巨大的机遇。



## 2. 技术篇





计算机视觉的内涵丰富，需要完成的任务众多，关键任务包括：图像增强、图像分类、图像检测与定位、图像分布、目标识别。本篇首先从计算机视觉的关键技术入手，再通过 AMiner 大数据平台挖掘获取了与领域相关的 66,519 项专利和 1,052 项国家基金支持情况，以此展览计算机视觉技术未来可能的发展方向。

## 2.1 图像增强

图像增强处理是数字图像处理技术中的一种重要方法。在实际生活中，图像可能会因拍摄环境恶劣、传输噪声引入等原因导致图像质量降低。图像增强处理可以有效去除图像噪声、增强图像边缘，突出图像中所需的重要信息，去除或弱化不重要的信息，达到改善图像的视觉质量的效果，更适合人的观察或机器的识别<sup>[15]</sup>。作为计算机视觉的重要组成部分，图像增强对于提升图像的质量发挥着重要的作用。

### 2.1.1 基于多尺度分析的图像增强

根据生理学家对人类视觉系统的研究结果，一种“最优”的图像表示方法应该具有多分辨率、局域性、方向性和各向异性的特性<sup>[16]</sup>，其基础的支撑区间应为“长条形”，能充分利用图像的几何特征，把这种“最优”的表示称为“多尺度几何分析”（Multiscale Geometric Analysis, MGA）。

图像的多尺度几何分析方法可分为自适应和非自适应两类。自适应方法是指图像变换的基函数随图像内容变化而变化，它一般先进行边缘检测，再利用边缘信息对原函数进行最优表示，主要有 Bandelet<sup>[17-18]</sup>、Beamlet<sup>[19]</sup>、Directionlet<sup>[20]</sup>、Terolelet<sup>[21]</sup>等。与之不同的是，非自适应方法是指图像变换的基函数与图像内容无关，它不需要先验地知道图像本身的几何特征，但是能对特定函数具有较为理想的逼近效果，其代表为 Ridgelet<sup>[22]</sup>、Curvelet<sup>[23]</sup>、Contourlet<sup>[24-26]</sup>、Shearlet<sup>[27]</sup>等。

### 2.1.2 数学形态增强

数学形态学<sup>[28]</sup>是图像处理和分析的新理论、新方法，它的基本原理是把图像看作一个集合，用“探针”（即某种形状的结构元素）对图像进行求补、移位、交或并的集合运算，这些集合运算就构成了各种不同的数学形态学方法。学者们通过引入数学形态学的知识，改进传统的图像增强方法，并运用于灰度图像增强、彩色图像增强和医学图像增强等情境下。

### 2.1.3 卷积神经网络增强

近年来,随着以卷积神经网络(CNN)的深度学习技术的发展,以图像去噪、图像超分辨和图像去模糊等为代表的图像复原与增强、以及以视觉生成与合成为代表的图像编辑问题都获得了较多的关注。2014年, Schmidt 等针对图像复原问题,提出了基于逐次迭代学习的判别学习模型——基于半二次分裂算法的级联收缩场(Cascade of Shrinkage Fields, CSF),该方法通过将预测过程展开为迭代学习算法,从训练数据中学习阶段模型参数<sup>[29]</sup>。2015年, Chen 等从反应扩散方程的角度出发,提出了非线性反应扩散(Trainable Nonlinear Reaction Diffusion, TNRD)模型,对每次迭代的滤波器和响应函数进行学习,并从递归神经网络的角度对模型进行了解释<sup>[30]</sup>。受 CSF 和 TNRD 启发, Zhang 等设计了一种基于卷积神经网络的深度去噪网络——DnCNN(Denoising Convolutional Neural Network)。该模型通过端到端的残差学习,从函数回归角度用卷积神经网络将噪声从噪声图像中分离出来,取得了显著优于其他方法的去噪结果<sup>[31]</sup>。

## 2.2 图像分类

作为计算机视觉领域的基础性任务,图像分类是目标检测、语义分割的重要支撑,其目标是将不同的图像划分到不同的类别,并实现最小的分类误差。经过近 30 年的研究,图像分类已经成功应用至社会生活的方方面面。如今,在我们的生活中随处可见,例如智能手机的相册自动分类、产品缺陷识别、无人驾驶等等。根据分类任务的目标不同,可以将图像分类任务划分成两部分:(1)单标签图像分类;(2)多标签图像分类。

### 2.2.1 单标签分类

单标签分类是简单的分类任务,图片的内容相对简单,只包含一个物体或者场景。单标签图像分类是指每张图片对应一个类别标签,根据物体类别的数量,又可以将单标签图像分类划分成二分类、多类别分类。

### 2.2.2 多标签分类

早期的图像分类方法主要对图像进行二分类或多类别分类,每次只需为图像选择一个类别标签,这是最常见的单标签分类方法。然而,在真实世界中,一幅图像往往不只包含单一的语义,因此,多标签学习方法在实际应用中更具有现实意义,这也更加符合人的认知习惯。多标签图像分类可以告知我们图像中是否同时包含这些内容,这也能够更好地解决实际生活中的问题。

## 2.3 图像检测与定位

物体检测为许多视觉任务提供了动力，例如实例分割<sup>[32]</sup>、姿势估计<sup>[33-35]</sup>、追踪<sup>[36]</sup>和动作识别<sup>[37]</sup>。它在监视<sup>[38]</sup>、自动驾驶<sup>[39]</sup>和视觉问答<sup>[40]</sup>中具有下游应用。物体检测器通过紧密围绕物体的轴对齐边界框来表示每个物体<sup>[41-46]</sup>。然后，他们将物体检测简化为具有大量潜在对象边界框的图像分类。对于每个边界框，分类器确定图像内容是特定的物体还是背景。单级检测器<sup>[44-45]</sup>在图像上滑动可能的边界框（称为锚点）的复杂排列，并在不指定框内容的情况下直接对其进行分类。两级检测器<sup>[41-42,44-46]</sup>重新计算每个电位的图像特征，然后对这些特征进行分类。

### 2.3.1 物体定位

目标检测的主要目的是从图片中检测并定位特定的多个目标。传统检测模型通常采用人工特征提取方法获得目标的特征描述，然后输入到一个分类器中学习分类规则。传统方法的弊端：一是人工提取特征方法复杂，并需要对检测目标有一定的先验知识；二是分离特征提取和分类训练，若特征提取不够好，训练就难以有好效果；三是高度依赖具体任务，可移植性差，一旦检测目标有较大变动，就要重新设计算法<sup>[47]</sup>。卷积神经网络通过卷积运算让计算机自动从图像中提取目标特征，这样获得的特征更自然，并且通用性好，对一定程度的扭曲形变有良好的鲁棒性，并且在图像分类上取得了巨大成功，使得人们开始研究它在计算机视觉其他领域的效果。

### 2.3.2 关键点检测

关键点检测是许多计算机视觉任务的基础，广泛应用于公共安全、智能人机交互、自动驾驶、步态识别等场景。基于深度学习的人体关键点检测虽然发展时间很短，但是发展迅猛，近几年涌现出很多优秀的关键点检测算法，比如 CPM<sup>[48]</sup>、SHN<sup>[49]</sup>、CPN<sup>[50]</sup>、RMPE<sup>[34]</sup>等等。

## 2.4 图像分割

对于一张图来说，图上可能有多个物体、多个人物甚至多层背景，希望能做到对于原图上的每个像素点都能预测它是属于哪个部分（人、动物、背景……）。

### 2.4.1 语义分割

图像语义分割的目标是对图像中每一个像素点进行类别预测，因此又称为密集像素点预测，语义分割网络的输入一般是 RGB 图像，输出也是图像，输出图

像中每个像素点的数值为类别编号。图像中属于同一类别的像素点会被划分为同一颜色表现，不同类别的像素被分割为不同颜色的区块，如图 4 所示，输入图像被分割为行人、自行车、背景三种类别，其中，行人类别使用棕红色表示，自行车类别使用绿色来表示，背景类别使用黑色表示。

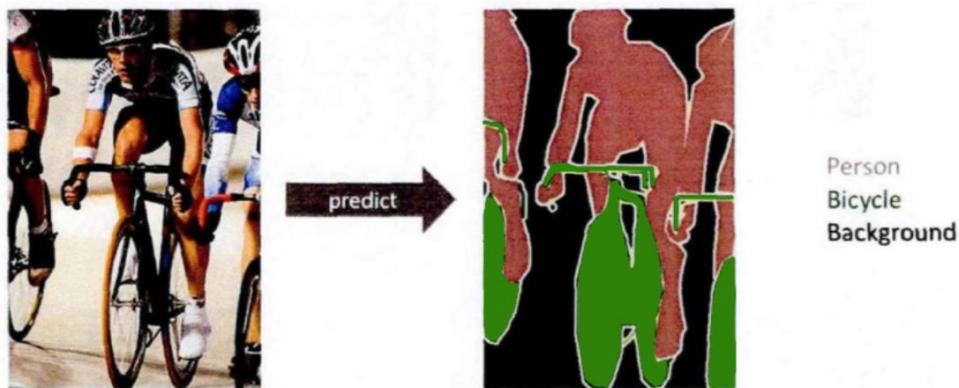


图 4 语义分割示例

#### 2.4.2 实例分割

实例分割（下图右下角）其实就是目标检测和语义分割的结合。相对目标检测的边界框，实例分割可精确到物体的边缘；相对语义分割，实例分割需要标注出图上同一物体的不同个体。如图 5 所示，在一张含有三只不同羊和一条狗的图片上，实例分割需要识别出不同羊的形状，而语义分割则无需分辨出不同羊的轮廓。相较于语义分割而言，实例分割更接近我们人类对世界的认知，而且允许对场景构成元素直接进行后续处理，如对行人进行动作识别等。实例分割综合了目标检测和语义分割等多个任务，实现起来也具有一定的挑战性，当前最优秀的实例级分割方法是深度卷积神经网络。

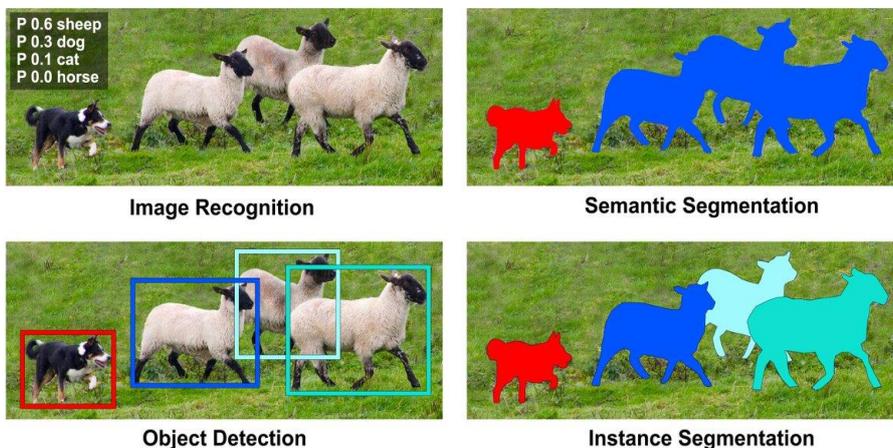


图 5 实例分割示例

### 2.4.3 全景分割

与之前介绍的语义分割与实例分割不同，全景分割任务（Panoptic Segmentation）要求图像中的每个像素点都必须被分配给一个语义标签和一个实例 id，如果无法确定可以给予空标注。其中，语义标签指的是物体的类别，而实例 id 则对应同类物体的不同编号。与语义分割相比，全景分割的困难在于要优化全连接网络的设计，使其网络结构能够区分不同类别的实例；而与实例分割相比，由于全景分割要求每个像素只能有一个类别和 id 标注，因此不能出现实例分割中的重叠现象。

## 2.5 目标识别

目标识别的目的在于判断场景（二维图像、视频或三维图像）中是否存在感兴趣目标，若存在则对其位置和姿态等信息进行估计<sup>[51]</sup>，它是计算机视觉中非常重要的一个研究方向。设计一个具有足够的通用性、稳健性，且简单实用的系统，能够在各种环境下无需太多约束和人类的干预的情况下自动对场景中的目标进行识别是目标识别研究的目的。目前目标识别在人类实际的生产和生活中具有非常广泛的应用和实用价值，例如目标跟踪、视频监控、信息安全、自动驾驶、图像检索、医学图像分析、无人机导航、遥感图像分析、国防系统等。

### 2.5.1 3D 目标识别

3D 目标识别，起初是对一些由简单的几何体组成的三维目标进行识别，例如柱体、立方体、椎体等。但是由于这些简单几何体的表达能力有限，现实世界中绝大多数物体很难由这些简单的几何体组合表示。因此过去的二十多年间，三维目标识别主要是对自由形状的三维目标进行识别<sup>[32]</sup>。自由形状是除了在边缘、顶点和拐角处外其余部分都有连续的法线构成的三维形状，现实世界中的绝大部分目标都可以认为是自由形状的目标，如人体、建筑物、雕塑、汽车等。目前的三维目标识别方法主要有：基于模型的方法；基于外观的方法；基于全局特征的方法。

- 基于模型的方法

该方法需要待识别物体模型的先验知识，建立模型库。该算法的流程如图 6 所示，首先为待识别的目标设计三维模型，并建立模型库；其次通过传感器获取待识别的真实目标的三维数据；然后对三维数据进行分析；最后进行模型匹配。

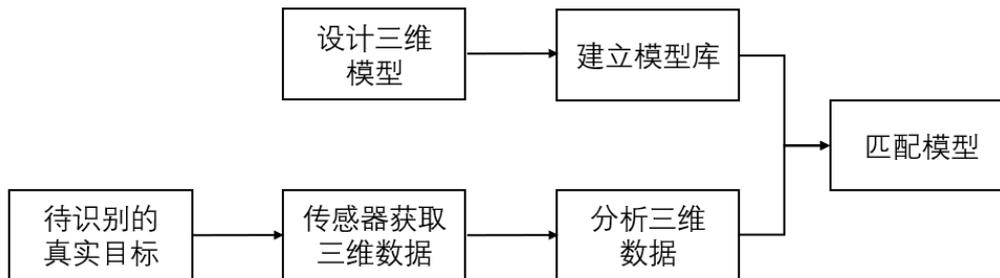


图 6 基于模型的 3D 目标识别方法流程

- 基于外观的方法

该方法是通过目标外观的相似性进行识别，其无需提前建立三维模型。该方法的步骤为：首先通过训练学习待识别的三维目标在二维图形中呈现的各种姿态；然后通过分析场景二维图形中各个物体的姿态，判断是否存在待识别的目标。

- 基于全局特征的方法

该方法在目标识别过程中将三维目标作为一个整体，提取整个三维目标的全局特征进行识别。这类方法速度非常快，在三维形状分类和检索中广泛应用，现有的典型方法有视点特征直方图<sup>[52]</sup>、三维几何距离<sup>[53]</sup>以及形状分布<sup>[32]</sup>等。

### 2.5.2 点云目标识别

随着计算机视觉技术的快速发展，基于三维点云数据的目标识别研究受到越来越广泛的关注。三维点云数据的目标识别一般包括特征表达和特征匹配策略两个部分，而匹配识别算法是关键组成部分，也是目前急需要攻克的难点。物体特征可以分为全局特征和局部特征，特征匹配的算法可以分为直接特征点匹配与间接特征点匹配方法。目前，基于三维点云数据的目标识别方法有多种，国内外许多专家学者对此做了大量的深入研究，并取得丰硕的成果。

上世纪八十年代中期，Besl 等提出一种三维形状的配准方法，称为最邻近迭代（Iterative Close Point, ICP）算法<sup>[54]</sup>。ICP 算法的主要思想是利用迭代的原理，找到刚性变换矩阵  $T$ ，使得场景中的点  $S$  和几何模型中的点  $M$  达到最优匹配。该算法可以准确有效地解决自由形态的点云配准问题，但其主要针对的是全局匹配，对于局部匹配需要多次平移和旋转，计算量显著增大，准确度也有所降低，效果不佳，且鲁棒性不好，对于有闭塞的情况更难以有效解决。Gregory C. Sharp 等提出改进的 ICP 方法，使用欧几里得几何学的不变特征结合 ICP 算法进行配准，称之为 ICPIF 算法<sup>[55]</sup>。该方法能够有效地解决深度图像配准难的问题，但计

算量依然比较大。A. Johnson 和 M. Hebert<sup>[56]</sup>提出一种“利用旋转图像有效识别杂乱三维场景下的目标”的方法。该方法提出使用旋转图像，模型用曲面网格表示，并由三维点转为二维空间点和参数表示的旋转图像。该方法可靠性好，但过程比较复杂。Mahmoudi 等提出将三维数据降到二维平面，通过三维特征和二维边界点的曲率分布特征实现目标识别<sup>[57]</sup>。

## 2.6 专利申请情况

根据专家推荐的“计算机视觉”领域关键词，从智慧芽专利数据库 (<https://analytics.zhuiyia.com/>) 中查找出全球 2000 至 2019 年期间“标题和摘要”中包含领域关键词的专利申请 66,519 件，具体如图 7 错误!未找到引用源。所示。

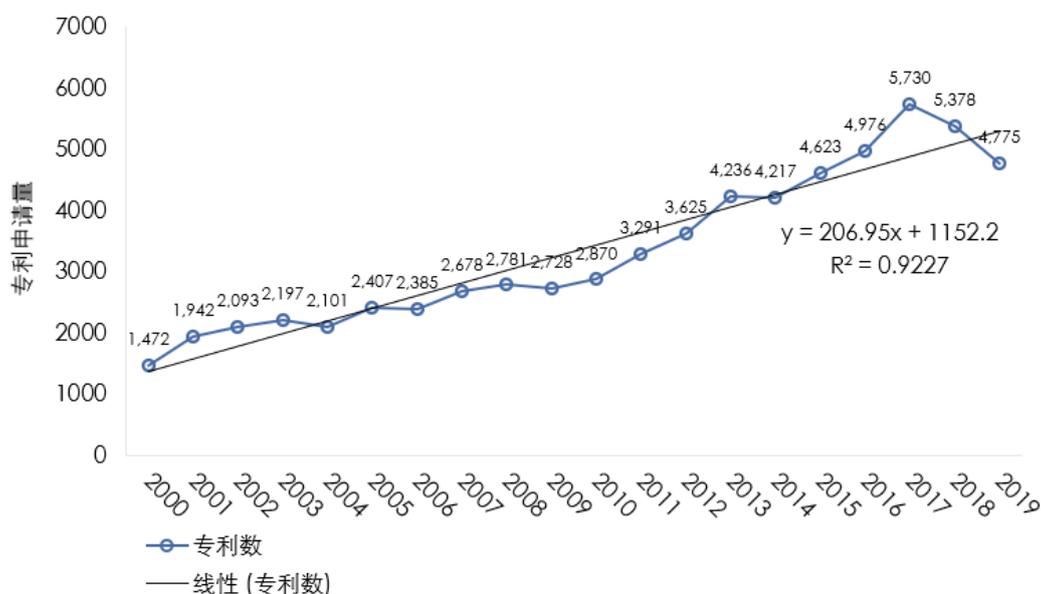


图 7 计算机视觉领域专利申请情况

过去 20 年，计算机视觉领域的专利申请量整体呈现上升态势，并在 2017 年达到顶峰，2018 年以后领域专利申请热度有所降低，申请量首次出现连续两年下滑。

### 2.6.1 全球专利申请概况

从拥有专利的国家排名来看，中国（20,830 件）、美国（18,692 件）和日本（5,945 件）是申请计算机视觉领域专利最多的三个国家（见图 8）。中国申请的领域专利数量最多，但是领先地位是在 2011 年以后才开始出现的，并在此之后一直处于领域专利申请数量的首位。

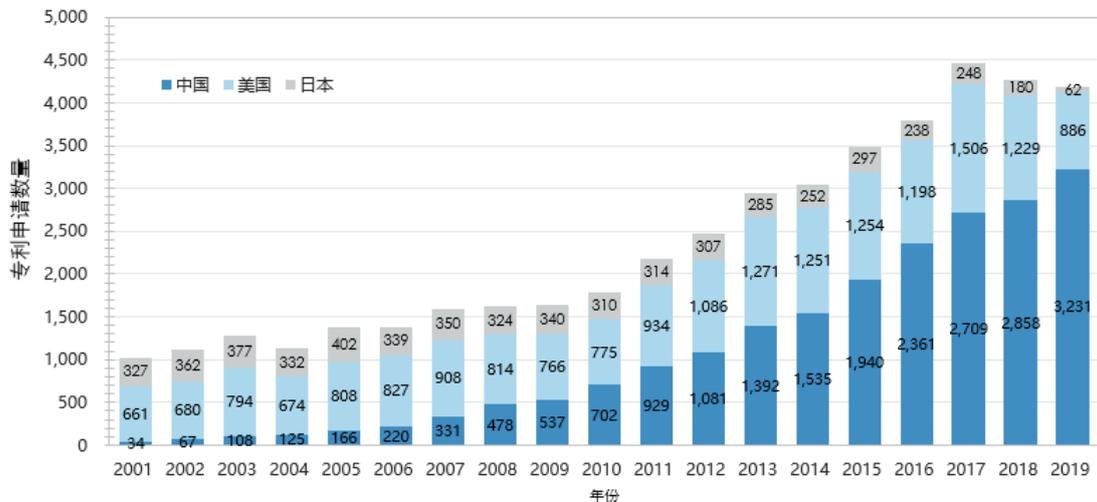


图 8 全球计算机视觉专利 TOP 3 国家年变化趋势

### 2.6.2 中国专利申请概况

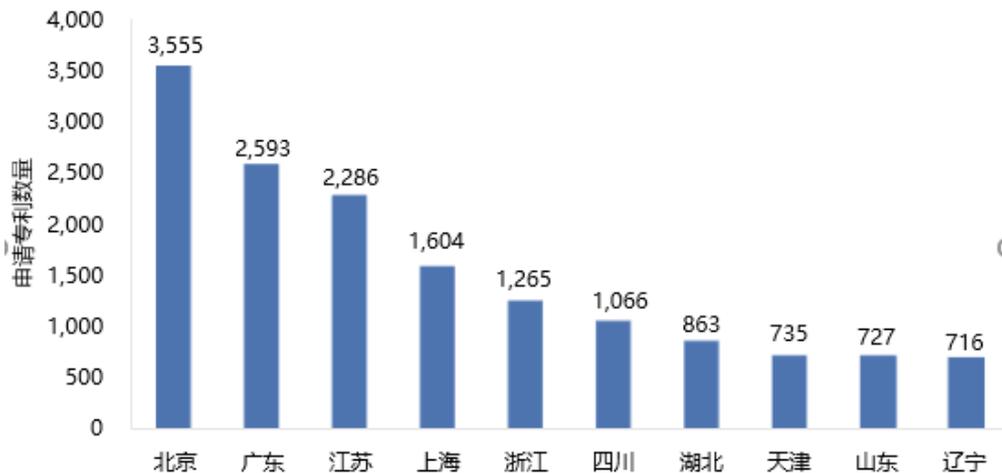


图 9 中国计算机视觉领域专利申请量 TOP 10 省市

从拥有专利的各省排名来看，北京、广东和江苏等经济发达省市的专利申请数量最多（见图 9）。图 10 是排名前三省市的专利年变化趋势，从图中可以看出，各省的专利申请数量都呈现稳定的上升态势，尤其是具备较强产业转化能力的广东省，在 2018 年以后甚至超越了人才高度聚集的北京市，成为申请计算机视觉领域专利申请量最多的省份。

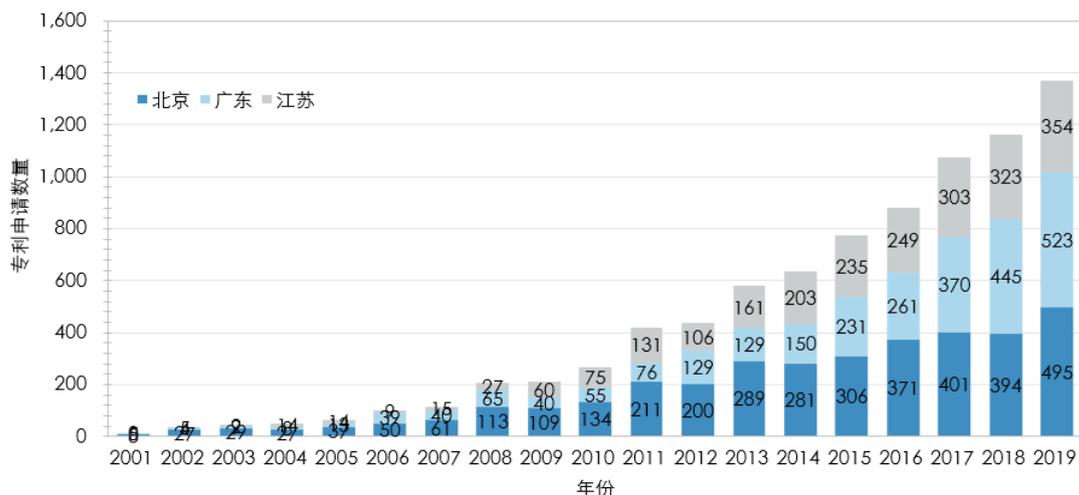


图 10 中国计算机视觉专利 TOP 3 省市年变化趋势

## 2.7 国家自然科学基金支持情况

根据“计算机视觉”领域关键词，从 AMiner 数据库中查找出 2010 至 2020 年国家自然科学基金支持的图数据库相关项目（包含未结题的项目）1,052 个，其中面上项目（451 个，42.87%）和青年科学基金项目（444 个，42.21%）的占比最高，具体分布如下所示。

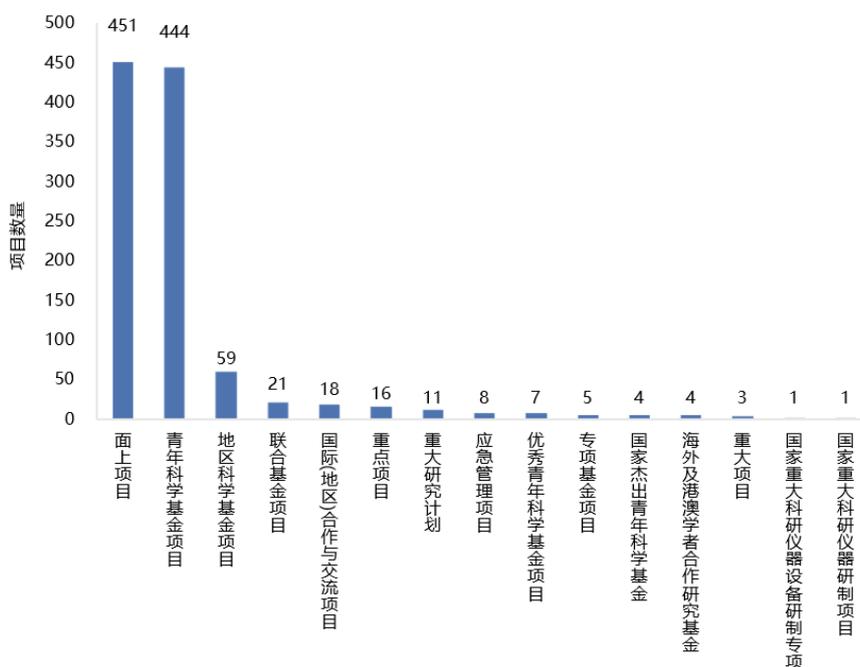


图 11 国家自然科学基金分布情况

国家自然科学基金项目数量前十的依托单位如图 12 所示。由图可见，中国科学院自动化研究所是项目批准量最多的机构，共计有 60 个，其中面上项目 27

个，青年科学基金项目 22 个。另外，电子科技大学的青年科技学者表现不俗，是其中唯一一个青年项目数量多于面上项目的机构，展现出较大的发展潜力。

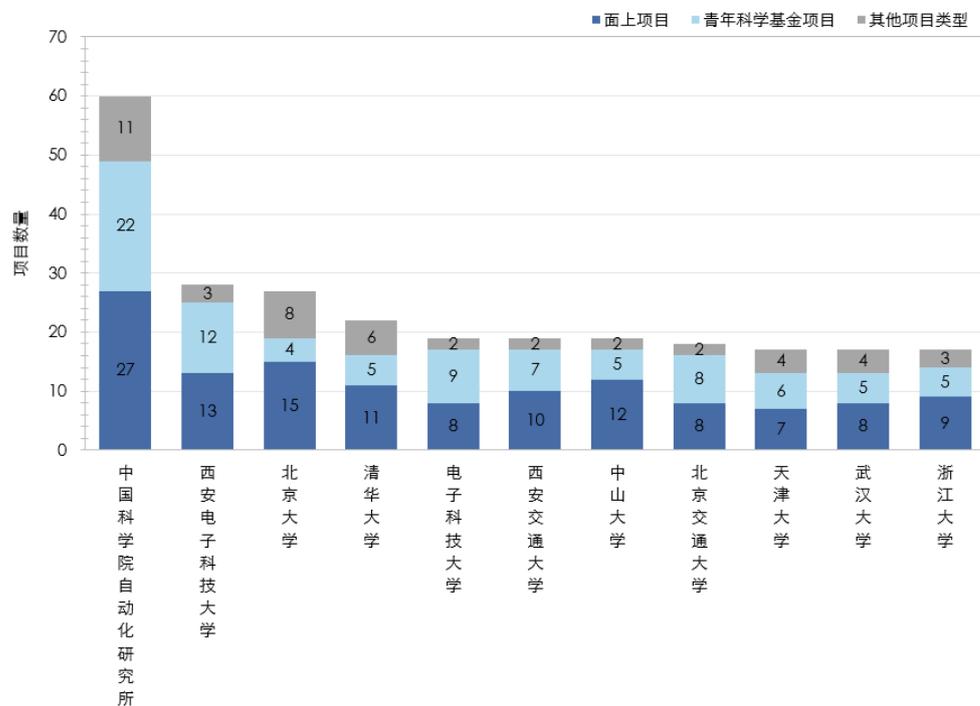


图 12 国家自然科学基金项目量前十的依托单位

# 3. 人才篇





在大数据时代，计算机视觉技术不断迭代更新，覆盖人群和应用场景逐渐扩大，计算机视觉领域的众多学者专家们也在不断探索与研究。本篇将对本领域学者的分布情况和代表性学者进行简要介绍。

计算机视觉领域学者筛选的具体方法如下：首先，通过 AMiner 大数据平台挖掘计算机视觉领域学术会议及期刊：IEEE 国际计算机视觉与模式识别会议（IEEE Conference on Computer Vision and Pattern Recognition, CVPR）、IEEE 国际计算机视觉大会（IEEE International Conference on Computer Vision, ICCV）、欧洲计算机视觉国际会议（European Conference on Computer Vision, ECCV）、IEEE 模式分析和机器智能学报（IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI）、国际计算机视觉杂志（International Journal on Computer Vision, IJCV）、IEEE 图像处理汇刊（IEEE Transactions on Image Processing, TIP）、计算机视觉和图像理解（Computer Vision and Image Understanding, CVIU）、模式识别（Pattern Recognition, PR）、模式识别快报（Pattern Recognition Letters, PRL）的近 20 年论文，提取论文中所有学者信息，以此分析学者的分布情况。从中选出与计算机视觉领域关键词相关学者 22,883 位，再按照学者的 h-index 进行排序，最后对其中排名靠前的部分学者进行简要介绍。领域关键词由计算机视觉顾问组给出，具体包括计算机视觉（Computer vision）、图像和视频采集（Image and video acquisition）、图像识别（Image recognition）、模式识别（Pattern recognition）、图像理解（Image understanding）、图像分类（Image classification）、目标检测（Object detection）、目标跟踪（Object tracking）、语义分割（Semantic segmentation）、实例分割（Instance segmentation）、卷积神经网络（Convolutional neural networks）、三维视觉（3D vision）、光流估计（Optical flow estimation）、景深估计（Scene depth estimation）、人脸识别（Face recognition）、图像生成（Image generation）。

## 3.1 计算机视觉学者概览

### 3.1.1 全球学者概况

#### ● 学者地图

学者分布地图对于计算机视觉领域学者调查、分析各地区竞争力现状尤为重要，图 13 为计算机视觉领域全球顶尖学者分布情况。其中，颜色越趋近于红色，表示学者越集中；颜色越趋近于绿色，表示学者越稀少。在全球范围内，计算机

视觉的顶尖学者主要分布于东亚、北美以及欧洲，此外，南美、澳大利亚等地亦有部分学者分布。



图 13 计算机视觉全球顶尖学者分布

### ● 国家对比

根据 AMiner 平台数据分析不同国家“计算机视觉”领域学者的数量，具体分析方法是根据论文作者的国家信息，将论文分类到各个国家中，从而统计出每个国家的学者数量。图 14 展示了领域学者数量前 10 的国家，由图可知，中国位居全球第一，随后为美国、英国、日本、德国等。

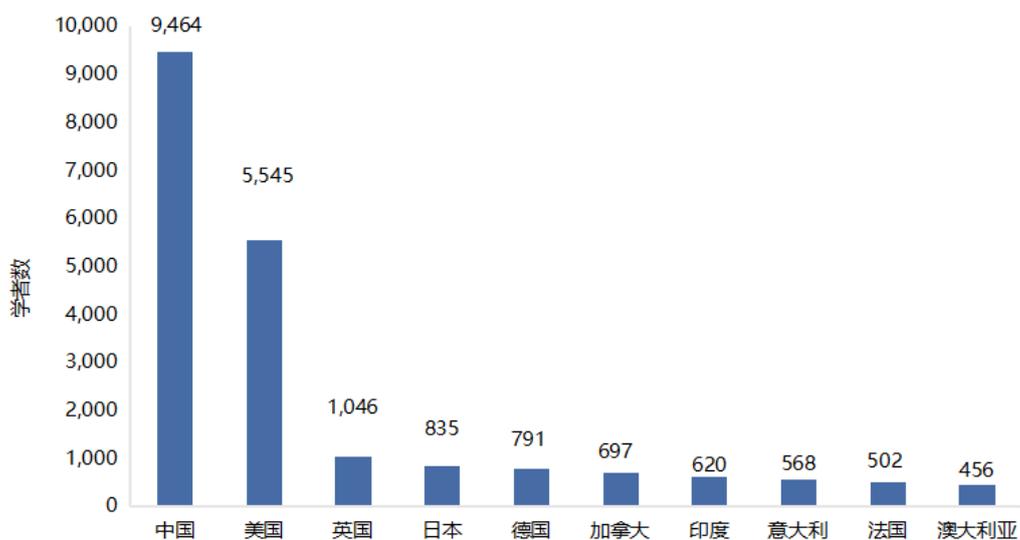


图 14 计算机视觉学者数量 Top 10 国家

### ● 学者 h-index 分布

**h-index** 是衡量一位学者影响力的重要指标之一，利用 AMiner 大数据平台统计计算机视觉领域排名前 2000 位学者的 **h-index**，具体分布如下图所示。其中，**h-index** 在 30~40 之间的人数最多为 1,413 人，占总学者数量的 70.65%；**h-index** 超过 60 的学者为 227 人，占总学者数量的 11.35%（图 15）。

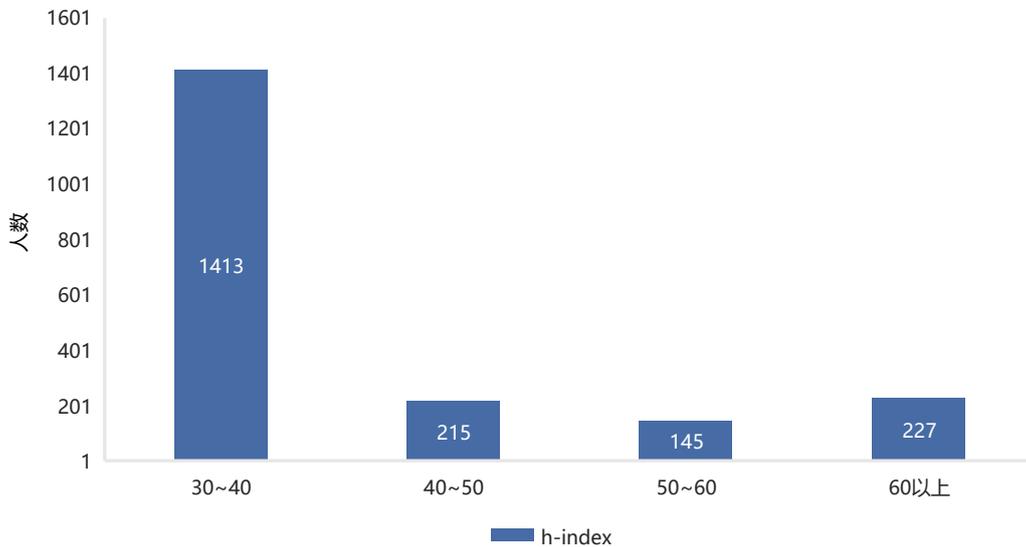


图 15 计算机视觉学者 h-index 分布

### ● 学者迁徙

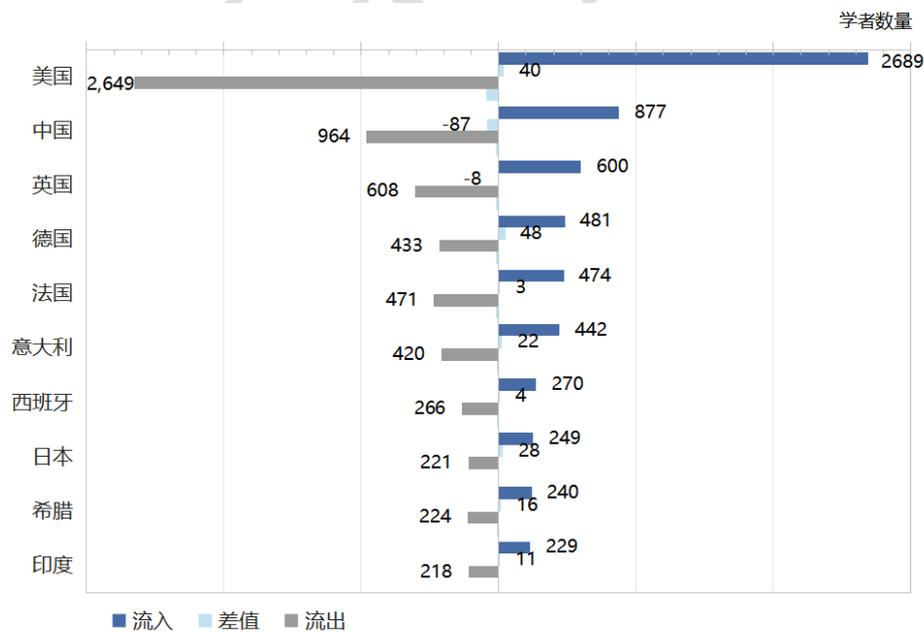


图 16 计算机视觉全球学者迁徙图

AMiner 可以对计算机视觉领域学者的迁徙量进行分析，如图 16 所示。全球计算机视觉领域流动总量最高的是美国，遥遥领先于排名第二的中国，随后为英

国、德国和法国等。此外，美国、德国、法国的人才流入大于人才流出，领域人才有所聚集；而中国和英国的人才流入量小于人才流出，出现领域人才流失现象。

### ● 机构对比

通过 AMiner 平台挖掘论文中的作者单位信息，将论文映射到各个单位机构中，统计每个机构的论文发表数量，并按照论文发表数量从高到低对机构进行了排序，如图 17 所示。

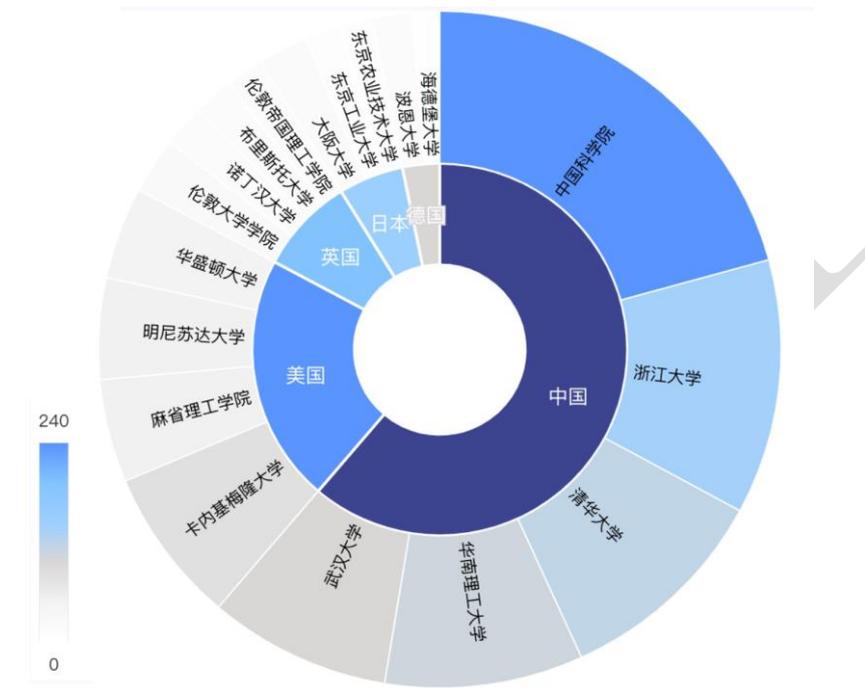


图 17 计算机视觉学术机构对比

从上图可以看出，中国、美国、日本、英国和德国拥有的顶尖科研机构最多。其中，中国科研机构发表的论文数量最多，排名前三的机构分别为中国科学院、浙江大学和清华大学。美国科研机构紧随其后，发表论文总数位居世界第二，其中排名前三的科研机构分别为卡内基梅隆大学、麻省理工学院和明尼苏达大学。

### 3.1.2 中国学者概况

#### ● 中国学者分布

中国计算机视觉领域人才分布如图 18 所示，领域学者大部分聚集于北京地区。与此同时，江苏、上海、辽宁、湖北、广东等地同样分布着可观数量的计算机视觉领域学者。北京地区学者主要聚集在以北大、清华、亚洲微软研究院为首的海淀人才汇集地，人才密度可谓在世界范围内都首屈一指，为该领域的发展提供了强大的人才保障。此外，香港、深圳、广州等城市在发展计算机视觉产业上

各具优势及特色，从研发成果产业化的角度看，香港有较顶尖的人才储备，而深圳有强大的科创队伍，广州则具备很强的产业化能力。

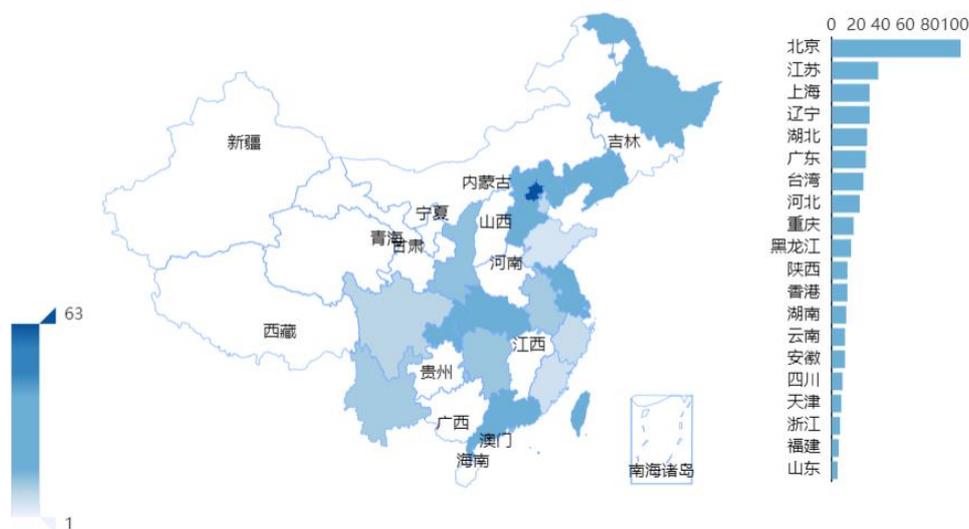


图 18 中国计算机视觉领域学者分布

### ● 中外合作情况

中国与其他国家在计算机视觉领域的合作情况可以由 AMiner 大数据平台分析得到，在上述期刊/会议中合作论文数量 TOP10 的关系如表 1 所示，论文数量是表现两国合作关系的重要指标，中美合作的论文数遥遥领先，体现出中美两国在本领域的密切合作关系。引用数量是衡量论文被认可度和质量的重要指标，中国与瑞士合作论文的平均引用数量最高，论文质量较高。

表 1 中外国家合作统计

序号	合作国家	论文数	平均引用数
1	美国	1321	156.1
2	新加坡	194	81.34
3	英国	122	67.73
4	澳大利亚	120	56.53
5	加拿大	103	94.62
6	日本	80	32.24
7	法国	40	125.4
8	印度	34	52.35
9	瑞士	29	196.71
10	德国	28	101.33

## 3.2 计算机视觉代表性学者介绍

AI 2000 人工智能全球最具影响力学者榜单（以下简称 AI 2000）由 AMiner 提供学术数据支持，涉及经典 AI（AAAI/IJCAI）、机器学习、计算机视觉、自然语言处理、机器人、知识工程、语音识别、数据挖掘、信息检索与推荐、数据

库、人机交互、计算机图形、多媒体、可视化、安全与隐私、计算机网络、计算机系统、计算理论、芯片技术和物联网等 20 个子领域，每个子领域排名靠前的 10 位研究者会进入榜单。如需获取更多 AI 2000 信息，请登录我们的网站 <https://www.aminer.cn/>。

参照榜单，我们收集整理了获得过 AI 2000 计算机视觉子领域的学者，通过“学者画像”的方式进行展示，另外还包括了学者的基本信息、AMiner 获奖信息、研究兴趣和相关代表性论文等。

2020 年 4 月 25 日，计算机视觉领域科学家 Thomas S. Huang（黄煦涛）院士逝世。为了纪念黄煦涛院士在计算机视觉领域作出的开创性研究贡献，我们诚意制作了黄煦涛院士的学者画像来回顾先生的学术生涯，以此缅怀。

### ●Trevor Darrell

教授

加州大学伯克利分校

2020 AI 2000 Most Influential Scholar Award in Multimedia

2020 AI 2000 Most Influential Scholar Award in Computer Vision

2018 AMiner Top 10 Most Influential Scholar Award in Multimedia

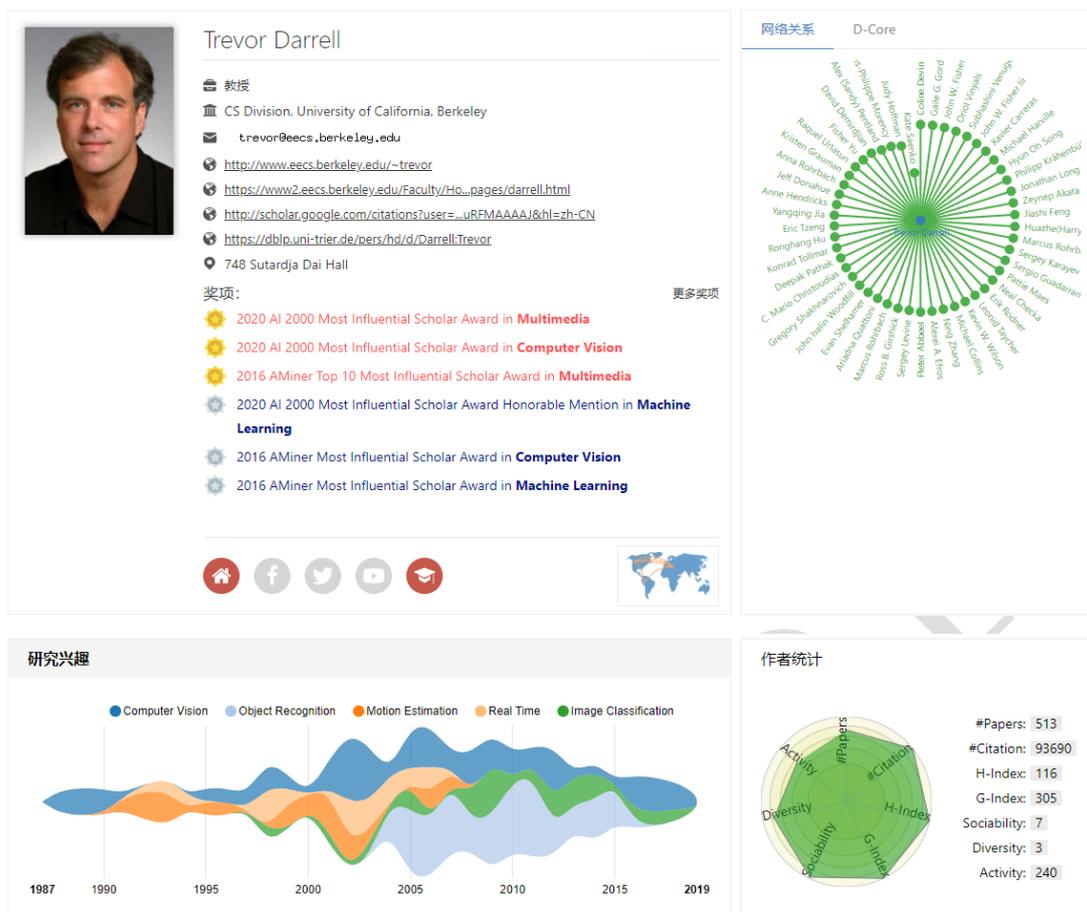
2018 AMiner Top 10 Most Influential Scholar Award in Computer Visio

2016 AMiner Top 10 Most Influential Scholar Award in Multimedia

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Machine Learning

2018 AMiner Most Influential Scholar Award in Machine Learning

2016 AMiner Most Influential Scholar Award in Computer Vision



Trevor Darrell，国际计算机科学研究所新成立的计算机视觉小组的负责人，目前在加州大学伯克利分校计算机科学系任教。他的兴趣包括计算机视觉、机器学习、计算机图形学和基于感知的人机界面。Trevor Darrell 于 1999 年至 2008 年在麻省理工学院 EECS 系任教，他在那里领导了视觉界面小组。1996 年至 1999 年，他是 Interval research Corporation 的研究人员，在 Ruzena Bajcsy 的 Grass lab 担任本科研究员，开始了计算机视觉的职业生涯。

### 相关论文精选

1. 论文标题: *Fully Convolutional Networks for Semantic Segmentation*

论文作者: Jonathan Long, Evan Shelhamer, **Trevor Darrell**

论文出处: *Computer Vision and Pattern Recognition*, (2015): 3431-3440

论文链接: <https://www.aminer.cn/pub/573697846e3b12023e66aa2d/>

2. 论文标题: *Caffe: Convolutional Architecture for Fast Feature Embedding*

论文作者: Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, **Trevor Darrell**

论文出处: *ACM Multimedia 2011*, (2014): 675-678

论文链接: <https://www.aminer.cn/pub/5550415c45ce0a409eb3a9a8/>

3. 论文标题: Long-term Recurrent Convolutional Networks for Visual Recognition and Description

论文作者: Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, **Trevor Darrell**

论文出处: IEEE Transactions on Pattern Analysis and Machine Intelligence,4(2016): 677-691

论文链接: <https://www.aminer.cn/pub/5550410f45ce0a409eb384d6/>

●**Ross B. Girshick**

研究科学家

脸书人工智能研究院 (Facebook AI Research, FAIR)

2020 AI 2000 Most Influential Scholar Award in Multimedia

2020 AI 2000 Most Influential Scholar Award in Computer Vision

2018 AMiner Top 10 Most Influential Scholar Award in Multimedia

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Machine Learning

2016 AMiner Most Influential Scholar Award in Multimedia

**Ross B. Girshick**

Research Scientist  
 Facebook AI Research  
 (617) 999 9687  
 rbg@eecs.berkeley.edu  
<https://www.rossgirshick.info/>  
<https://scholar.google.com/hk/citations?h...&user=W8VIEZgAAAAJ>  
[https://dblp.uni-trier.de/pers/hd/g/Girshick:Ross\\_B](https://dblp.uni-trier.de/pers/hd/g/Girshick:Ross_B)  
 Sutardja Dai Hall #266 Berkeley CA 94720-1774

奖项:

- 2020 AI 2000 Most Influential Scholar Award in **Multimedia**
- 2020 AI 2000 Most Influential Scholar Award in **Computer Vision**
- 2020 AI 2000 Most Influential Scholar Award Honorable Mention in **Machine Learning**
- 2016 AMiner Most Influential Scholar Award in **Multimedia**

网络关系 D-Core

Network graph showing connections to other researchers like Sergio Iqbal, Piotr F. Felzenszwalb, and others.

**研究兴趣**

Object Detection, Deformable Part Mode..., Object Recognition, Feature Extraction, Rgb-d Images

2006 2008 2010 2012 2014 2016 2018 2019

**作者统计**

#Papers: 94  
 #Citation: 98525  
 H-Index: 58  
 G-Index: 94  
 Sociability: 5  
 Diversity: 2  
 Activity: 169

Ross Girshick, 供职于 Facebook 人工智能研究院, 致力于计算机视觉和机器学习。2012 年, 他在 Pedro Felzenszwalb 的指导下获得了芝加哥大学计算机科学博士学位。在加入 FAIR 之前, Ross 是微软研究院的研究员, 也是加州大学伯克利分校 (University of California, Berkeley) 的博士后。他的兴趣包括实例级的对象理解和将自然语言处理与计算机视觉结合起来的视觉推理挑战。2017 年, 他获得过 PAMI Young 研究员奖, 并以开发 R-CNN (基于区域的卷积神经网络) 目标检测方法而闻名。2017 年, Ross Girshick 还因 “Mask R-CNN” 获得了 ICCV MARR 奖。

### 相关论文精选

1. 论文标题: *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*

论文作者: Shaoqing Ren, Kaiming He, **Ross B. Girshick**, Jian Sun

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., 6 (2017): 1137-1149

论文链接: <https://www.aminer.cn/pub/5736986b6e3b12023e730129/>

2. 论文标题: *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*

论文作者: **Ross B. Girshick**, Jeff Donahue, Trevor Darrell, Jitendra Malik

论文出处: CVPR '14 Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition: 580-587

论文链接: <https://www.aminer.cn/pub/556f622a2401b4b38c23635c/>

3. 论文标题: *Caffe: Convolutional Architecture for Fast Feature Embedding*

论文作者: Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, **Ross B. Girshick**, Sergio Guadarrama, Trevor Darrell

论文出处: ACM Multimedia 2001, (2014): 675-678

论文链接: <https://www.aminer.cn/pub/5550415c45ce0a409eb3a9a8/>

### ●Christian Szegedy

研究科学家

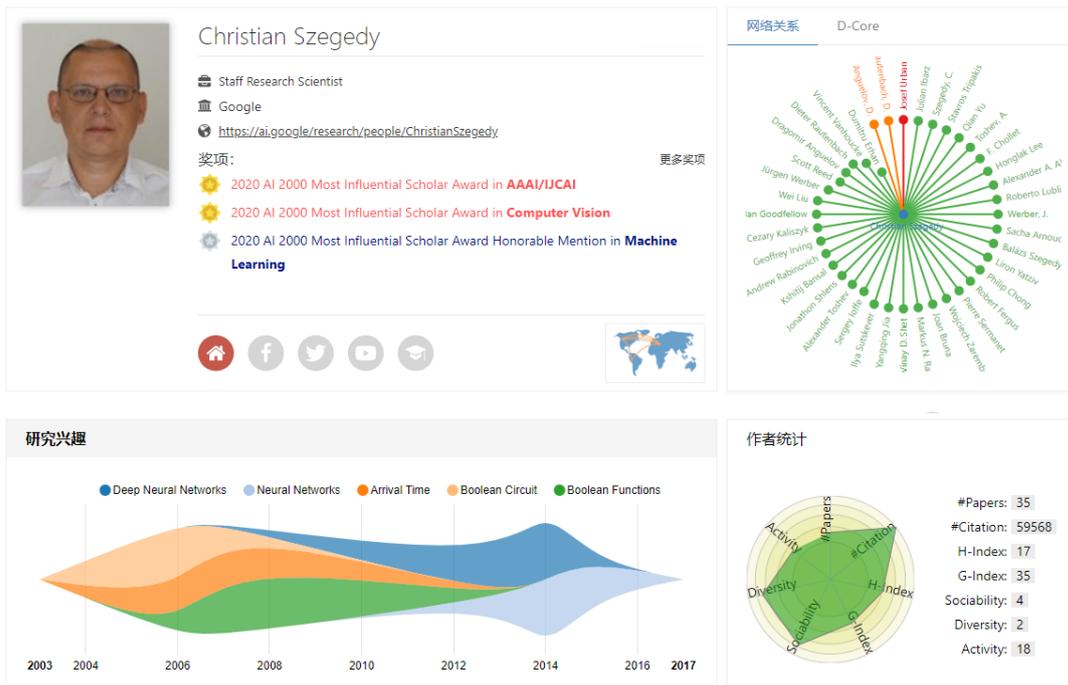
谷歌

2020 AI 2000 Most Influential Scholar Award in AAAI/IJCAI

2020 AI 2000 Most Influential Scholar Award in Computer Vision

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Machine

Learning



ChristianSzegedy, 波恩大学的数学博士, 曾在伯克利凯登斯研究实验室担任研究科学家, 研究领域有数字电路设计自动化、物理设计和逻辑综合。现在 Christian 是 Google 的一名研究人员, 通过深度学习研究机器学习、人工智能和计算机视觉。

相关论文精选

1. 论文标题: *Going Deeper with Convolutions*

论文作者: **Christian Szegedy**, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich  
 论文出处: Computer Vision and Pattern Recognition, (2015)

论文链接: <https://www.aminer.cn/pub/5550417d45ce0a409eb3bc08/>

2. 论文标题: *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*

论文作者: Sergey Ioffe, **Christian Szegedy**

论文出处: International Conference on Machine Learning, (2015)

论文链接: <https://www.aminer.cn/pub/573696ce6e3b12023e5ce95a/>

3. 论文标题: *Rethinking the Inception Architecture for Computer Vision*

论文作者: Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna

论文出处: CVPR, (2016)

论文链接: <https://www.aminer.cn/pub/5736960e6e3b12023e520c34/>

### ●Feifei Li (李飞飞)

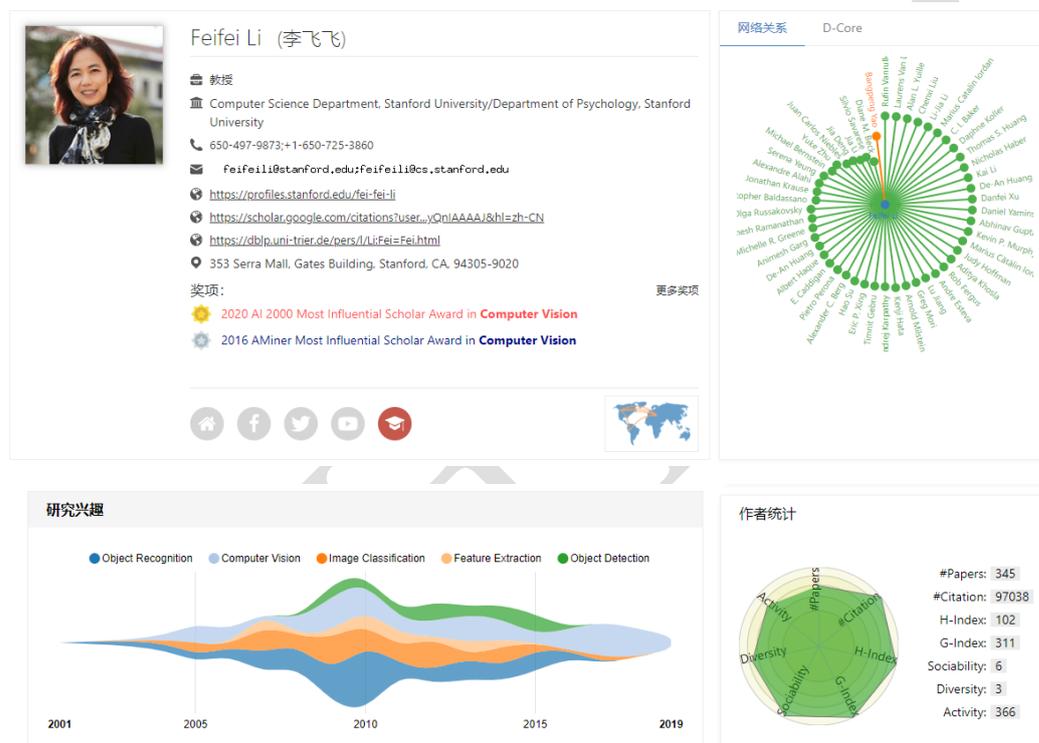
美国斯坦福大学红杉讲席教授

美国国家工程院院士

Twitter 公司董事会独立董事

2020 AI 2000 Most Influential Scholar Award in Computer Vision

2016 AMiner Most Influential Scholar Award in Computer Vision



李飞飞, 1976 年出生于北京, 长在四川, 16 岁随父母移居美国新泽西州。2015 年 12 月 1 日, 入选 2015 年“全球百大思想者”。2018 年 3 月, 获“影响世界华人大奖”。现为美国斯坦福大学红杉讲席教授, 美国国家工程院院士, 以人为本人工智能研究院 (HAI) 院长, AI4ALL 联合创始人及主席, Twitter 公司董事会独立董事。

李飞飞目前的研究兴趣包括认知性人工智能、机器学习、深度学习、计算机视觉和人工智能+医疗保健, 尤其是用于医疗保健的环境智能系统。在过去, 她还从事认知和计算神经科学。李飞飞在顶级期刊和会议上发表了 200 多篇科学文章, 包括自然杂志、PNAS、神经科学杂志、CVPR、ICCV、NIPS、ECCV、ICRA、

IROS、RSS、IJCV、IEEE-PAMI、新英格兰医学杂志。她是全国非营利 AI4ALL 的联合创始人和主席，旨在增加 AI 教育的包容性和多样性。

#### 相关论文精选

1. 论文标题: *Imagenet: A Large-Scale Hierarchical Image Database*

论文作者: Jia Deng, Wei Dong, Richard Socher, Li-jia Li, Kai Li, **Fei-fei Li**

论文出处: CVPR, (2009): 248-255

论文链接: <https://www.aminer.cn/pub/53e9b844b7602d970440513c/>

2. 论文标题: *ImageNet Large Scale Visual Recognition Challenge*

论文作者: Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, **Li Fei-Fei**

论文出处: International Journal of Computer Vision, (2015)

论文链接: <https://www.aminer.cn/pub/5550415945ce0a409eb3a820/>

3. 论文标题: *Large-Scale Video Classification with Convolutional Neural Networks*

论文作者: Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, **Li Fei-Fei**

论文出处: Computer Vision and Pattern Recognition, (2014): 1725-1732

论文链接: <https://www.aminer.cn/pub/5550460645ce0a409eb5b6e2/>

#### ● **Xiaoou Tang** (汤晓鸥)

教授

香港中文大学

商汤科技创始人

2020 AI 2000 Most Influential Scholar Award in Computer Vision

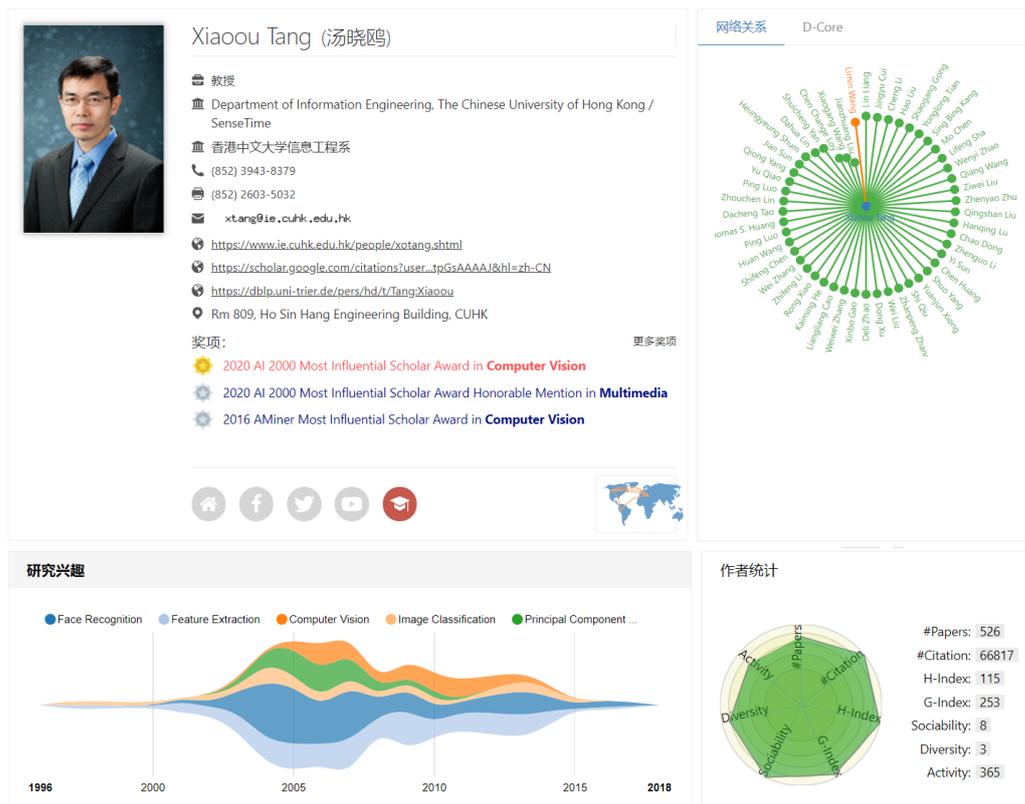
2018 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Multimedia

2018 AMiner Most Influential Scholar Award in Multimedia

2018 AMiner Most Influential Scholar Award in Machine Learning

2016 AMiner Most Influential Scholar Award in Computer Vision



汤晓鸥，1990 年于中国科学技术大学获得学士学位，1991 年于美国罗切斯特大学获得硕士学位，1996 年于麻省理工学院（MIT）获得博士学位，于香港中文大学信息工程系任教授。2005-2007 年于微软亚洲研究院，担任视觉计算组主任。2008 年 12 月起在深圳先进技术研究院多媒体集成技术研究室任主任和研究员。2014 年成立人工智能公司商汤科技，现任香港中文大学信息工程系教授、工程学院杰出学人，兼任中国科学院深圳先进技术研究院副院长。

汤晓鸥的研究领域包括多媒体、计算机视觉、模式识别及视频处理，已在这些领域发表论文 200 余篇。同微软合作已申请近 40 项专利。同微软亚洲研究院合作研发的图像识别技术已被用于微软图像搜索引擎，成为世界上第一项图像识别技术，被用于大规模商业应用。汤晓鸥发明的人脸识别技术，是世界上第一个超过人眼识别能力的计算机算法。其在香港中文大学创办的多媒体实验室，2016 年与麻省理工、斯坦福等著名大学一道，入选世界十大人工智能先锋实验室。而且为促进深港两地的研究人员交流互通和协同创新，汤晓鸥牵头组建了香港中文大学和中国科学院深圳先进技术研究院的联合实验室，并积极招募了一批海外优秀科研人才加入团队，继续在源头创新的道路上不断前行。2019 年 7 月，马来

西亚总理马哈蒂尔任命商汤科技创始人汤晓鸥担任马来西亚国家主权基金国库控股公司董事。汤晓鸥是马来西亚历史上首个担任此职位的外国人。

### 相关论文精选

1. 论文标题: *Single Image Haze Removal Using Dark Channel Prior*

论文作者: Kaiming He, Jian Sun, **Xiaoou Tang**

论文出处: *Single Image Haze Removal Using Dark Channel Prior*, (2011)

论文链接: <https://www.aminer.cn/pub/53e9abdab7602d970358f7c0/>

2. 论文标题: *Deep Learning Face Attributes in the Wild*

论文作者: Ziwei Liu, Ping Luo, Xiaogang Wang, **Xiaoou Tang**

论文出处: *International Conference on Computer Vision*, (2014)

论文链接: <https://www.aminer.cn/pub/5550411745ce0a409eb38760/>

3. 论文标题: *Learning to Detect A Salient Object*

论文作者: Tie Liu, Jian Sun, Nanning Zheng, **Xiaoou Tang**, Heung-yeung Shum

论文出处: *CVPR*, 2 (2011): 353-367

论文链接: <https://www.aminer.cn/pub/53e99bd5b7602d9702480674/>

### ●Jian Sun (孙剑)

教授

西安交通大学

旷视科技首席科学家

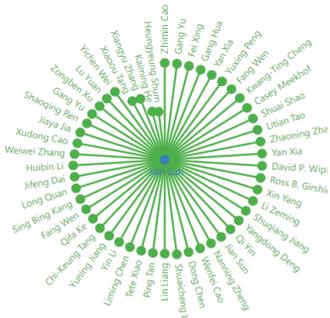


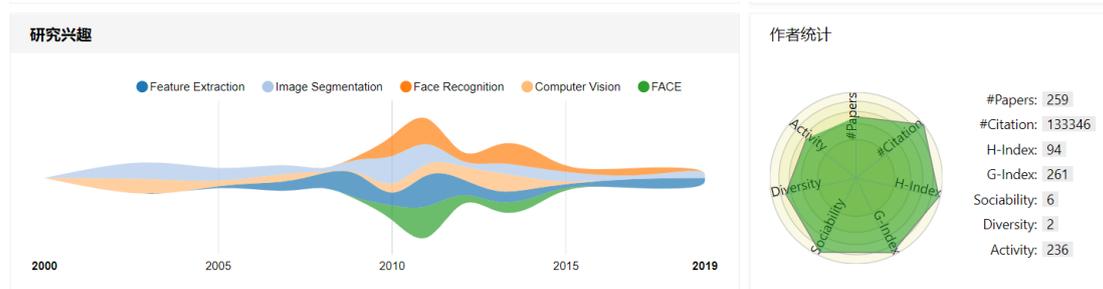
**Jian Sun (孙剑)**

其他  
Megvii Technology  
sun.jian@megvii.com  
<http://www.jiansun.org/>  
<https://scholar.google.com/citations?hl=en&rks&sortby=pubdate>  
[https://dblp.org/pers/hd/s/Sun\\_0001:Jian](https://dblp.org/pers/hd/s/Sun_0001:Jian)  
3F Tower A, Raycom InfoTech Park No.2 Kexueyuan South Road Haidian District, Beijing, China 100190

奖项: [更多奖项](#)  
2020 AI 2000 Most Influential Scholar Award in **Computer Vision**  
2020 AI 2000 Most Influential Scholar Award Honorable Mention in **Machine Learning**

网络关系 D-Core





孙剑，男，前微软亚研院首席研究员，现就职于北京旷视科技有限公司，任旷视首席科学家、旷视研究院院长。孙剑于 1997、2000、2003 年在西安交通大学自动控制专业获得学士、西安交通大学人工智能与机器人研究所获得硕士和博士。毕业后加入微软亚洲研究院，任首席研究员；2015 年至 2016 年在微软美国研究院任合伙人级研究主管。其主要研究方向是计算机视觉和深度学习。

孙剑自 2002 年以来在 CVPR、ICCV、SIGGRAPH、PAMI 等顶级学术会议和期刊上发表学术论文 100 余篇，两次获得 CVPR 最佳论文奖(2009, 2016)，两次获得 CVPR Best Paper Award。孙剑拥有超过 40 项美国或国际专利，于 2010 年被美国权威技术期刊 MIT Technology Review 评选为“全球 35 岁以下杰出青年创新者 TR35”。

### 相关论文精选

1. 论文标题: *Image Super-resolution Using Gradient Profile Prior*

论文作者: Kaiming He, Xiangyu Zhang, Shaoqing Ren, **Jian Sun**

论文出处: CVPR, (2008): 1-8

论文链接: <https://www.aminer.cn/pub/53e9ad05b7602d97036e4420/>

2. 论文标题: *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*

论文作者: Shaoqing Ren, Kaiming He, Ross B. Girshick, **Jian Sun**

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., 6 (2017): 1137-1149

论文链接: <https://www.aminer.cn/pub/5736986b6e3b12023e730129/>

3. 论文标题: *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*

论文作者: He Kaiming, Zhang Xiangyu, Ren Shaoqing, **Sun Jian**

论文出处: CVPR, (2007): 1-8

论文链接: <https://www.aminer.cn/pub/573696f46e3b12023e5f0d4d/>

●Kaiming He (何恺明)

研究员

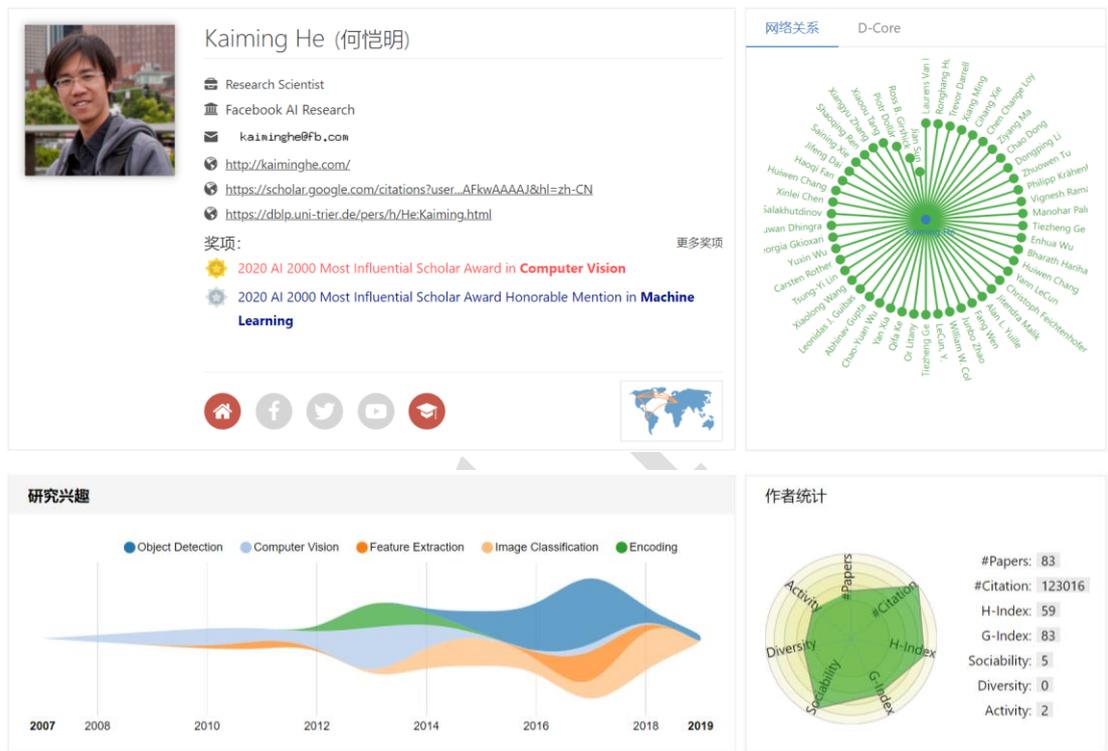
脸书人工智能研究院 Facebook AI Research (FAIR)

2020 AI 2000 Most Influential Scholar Award in Computer Vision

2018 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Machine Learning

2018 AMiner Most Influential Scholar Award in Machine Learning



何恺明，本科就读于清华大学，博士毕业于香港中文大学多媒体实验室。2011年加入微软亚洲研究院（MSRA）工作，主要研究计算机视觉和深度学习。2016年，加入 Facebook AI Research（FAIR）担任研究科学家。

何恺明的研究兴趣是计算机视觉和深度学习。2009年，何恺明成为首获计算机视觉领域三大国际会议之一 CVPR “最佳论文奖”的中国学者。在 2015 年的 ImageNet 图像识别大赛中，何恺明和他的团队用“图像识别深度差残学习”系统，击败谷歌、英特尔、高通等业界团队，荣获第一。何恺明作为第一作者获得了 CVPR 2009，CVPR 2016 和 ICCV 2017（Marr Prize）的最佳论文奖，并获得了 ICCV 2017 最佳学生论文奖。2017 年 4 月，何恺明获选香港中文大学工程学院杰出校友。2018 年，第 31 届计算机视觉和模式识别大会（Conference on

Computer Vision and Pattern Recognition, CVPR) 在美国盐湖城召开, 何恺明获得本届大会的 PAMI 年轻学者奖。

#### 相关论文精选

1. 论文标题: *Deep Residual Learning for Image Recognition*

论文作者: **Kaiming He**, Xiangyu Zhang, Shaoqing Ren, Jian Sun

论文出处: CVPR, (2016)

论文链接: <https://www.aminer.cn/pub/573696026e3b12023e515eec/>

2. 论文标题: *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*

论文作者: Shaoqing Ren, **Kaiming He**, Ross B. Girshick, Jian Sun

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., 6 (2017): 1137-1149

论文链接: <https://www.aminer.cn/pub/5736986b6e3b12023e730129/>

3. 论文标题: *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*

论文作者: Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

论文出处: ICCV, (2015)

论文链接: <https://www.aminer.cn/pub/573696f46e3b12023e5f0d4d/>

#### ●Jitendra Malik

教授

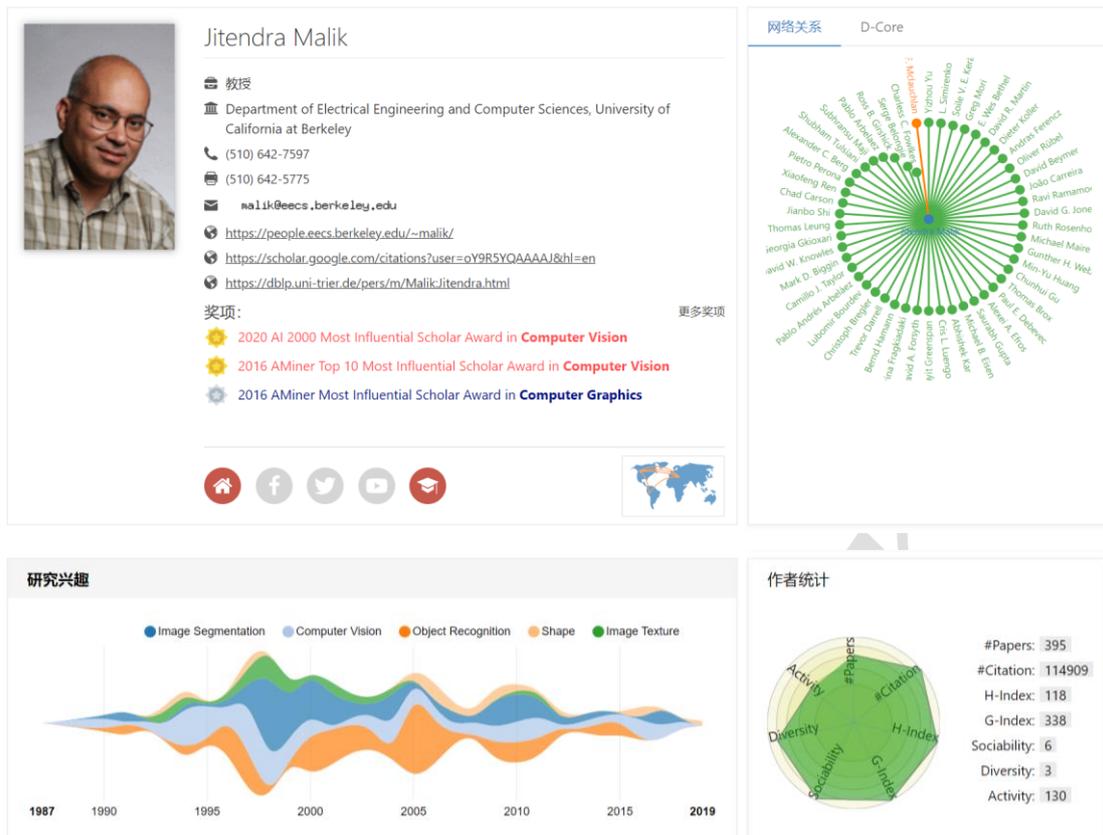
加州大学伯克利分校

2020 AI 2000 Most Influential Scholar Award in Computer Vision

2018 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2016 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2016 AMiner Most Influential Scholar Award in Computer Graphics



Jitendra Malik，加州大学伯克利分校（UC Berkeley）电子工程和计算机科学的教授，自 1986 年起就在该校任教。他还在生物工程系和认知科学与视觉科学组任职。

Jitendra Malik 的研究小组在计算机视觉、人眼视觉的计算建模、计算机图形学和生物图像分析等许多不同的领域开展了工作，共发表了 150 多篇研究论文和 30 多篇博士论文。在这项研究中出现了一些著名的概念和算法，如各向异性扩散、标准化切割、高动态范围成像和形状背景。

Jitendra Malik 于 1980 年获得电气工程最佳毕业生金奖，并于 2008 年获得印度工业技术学院坎普尔分校颁发的杰出校友奖。他于 1985 年获得斯坦福大学博士学位，1989 年获得总统青年调查员奖，在加州大学伯克利分校，他被选为 2000 年戴安麦肯太尔优秀教学奖和 2001 年米勒研究教授。他是 ACM 和 IEEE 的成员，也是美国国家工程院的成员。

●相关论文精选

1. 论文标题: *Normalized Cuts and Image Segmentation*

论文作者: Jianbo Shi, **Jitendra Malik**

论文出处: IEEE Transactions on Pattern Analysis and Machine Intelligence, 8 (2000): 731-737

论文链接: <https://www.aminer.cn/pub/53e9afe8b7602d9703a3af5a/>

2 论文标题: *Scale-Space and Edge Detection Using Anisotropic Diffusion*

论文作者: Pietro Perona, **Jitendra Malik**

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., 7 (1990): 629-639

论文链接: <https://www.aminer.cn/pub/53e9a6a6b7602d9702fdd057/>

3. 论文标题: *Shape Matching and Object Recognition Using Shape Contexts*

论文作者: S. Belongie, J. Malik, **J. Puzicha**

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., 4 (2002): 509-522

论文链接: <https://www.aminer.cn/pub/53e9a6a6b7602d9702fdd057/>

### ●Andrew Zisserman

教授

牛津大学

2018 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2016 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Computer Vision

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Machine Learning

2018 AMiner Most Influential Scholar Award in Machine Learning



**Andrew Zisserman**

教授

Computer Vision Engineering, University of Oxford

removethisifyouarehuman-az@robots.ox.ac.uk

<https://www.robots.ox.ac.uk/~az/>

<https://www.research.ox.ac.uk/Researchers/andrew-zisserman>

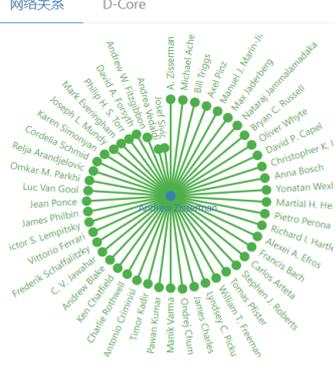
<https://scholar.google.co.uk/citations?hl=&user=UZ5wscMAAAAJ>

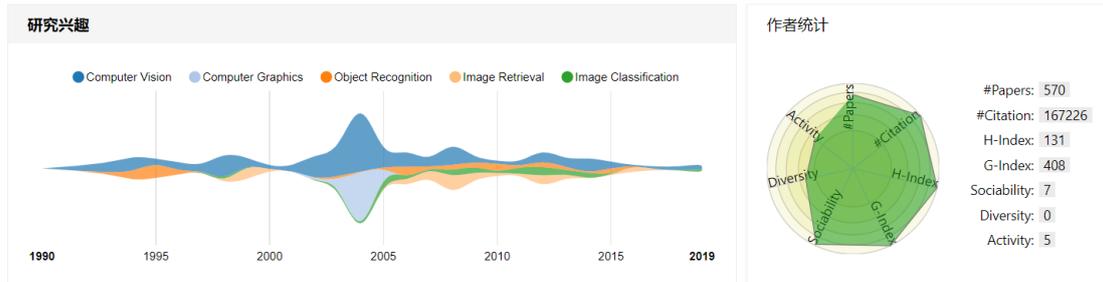
<https://dblp.uni-trier.de/pers/z/Zisserman/Andrew>

奖项: 更多奖项

- 2016 AMiner Top 10 Most Influential Scholar Award in **Computer Vision**
- 2020 AI 2000 Most Influential Scholar Award Honorable Mention in **Computer Vision**
- 2020 AI 2000 Most Influential Scholar Award Honorable Mention in **Machine Learning**

网络关系 D-Core





Andrew Zisserman, 英国牛津大学的视觉几何小组负责人。Andrew Zisserman 的研究兴趣包括视觉识别、图像检索、多视图几何和计算机视觉的其他方面。他是这一领域中论文被引用最多的学者之一，曾在顶级计算机视觉会议上获得多个奖项，例如国际计算机视觉会议上。他出版了几本书，包括《视觉重建》和《计算机视觉中的多视图几何》。同时 Andrew Zisserman 也是皇家学会的会员。

### 相关论文精选

1. 论文标题: *Multiple View Geometry in Computer Vision*

论文作者: Richard Hartley, **Andrew Zisserman**

论文出处: *Optics and Lasers in Engineering*, 1 (2002): 85-86

论文链接: <https://www.aminer.cn/pub/53e9a44fb7602d9702d6a959/>

2. 论文标题: *The Pascal Visual Object Classes (VOC) Challenge*

论文作者: Everingham Mark, Gool Luc, Williams K. Christopher, Winn John, **Zisserman Andrew**

论文出处: *International Journal of Computer Vision*, 303-338, 2010.

论文链接: <https://www.aminer.cn/pub/53e99fc2b7602d9702899ee6/>

3. 论文标题: *Video Google: A Text Retrieval Approach to Object Matching in Videos*

论文作者: Sivic Josef, **Zisserman Andrew**

论文出处: *ICCV*, (2003): 1470-1470

论文链接: <https://www.aminer.cn/pub/53e9b641b7602d970419b955/>

### ●Songchun Zhu (朱松纯)

教授

加州大学洛杉矶分校

2016 AMiner Most Influential Scholar Award in Computer Vision



朱松纯，1968 年出生于湖北省鄂州市，全球著名计算机视觉专家，统计与应用数学家、人工智能专家，现任美国洛杉矶加州大学统计系与计算机系教授，UCLA 计算机视觉、认知、学习与自主机器人中心主任。

朱松纯于 1996 年获美国哈佛大学计算机博士学位，师从国际数学大师大卫·曼福德教授，在国际顶级期刊和会议上发表论文 200 余篇，三次问鼎计算机视觉领域国际最高奖项——马尔奖。朱松纯在 1990 年代率先将概率统计建模与随机计算方法引入计算机视觉研究，提出了一系列图像与视频的结构化解译的框架、数理模型和统计算法，发展了广义模式理论（General Pattern Theory）。在认知科学领域，如视觉常识推理、场景理解等领域做出重要贡献。自 2010 年以来，朱松纯两次担任美国视觉、认知科学、AI 领域跨学科合作项目 MURI 负责人。朱松纯教授在科研方面具有很强的前瞻性，选题和方法独树一帜，长期致力于构建计算机视觉、认知科学、乃至人工智能科学的统一数理框架。

2005 年，朱松纯联合沈向洋等多位知名科学家在湖北省鄂州市创建民办、非营利性国际交流平台莲花山研究院，并任院长。研究院连续 5 年举办国际学术研讨会和暑期免费讲习班，为国内年轻学者和大量学生提供了一个学术氛围浓厚、具有国际科研水准的开放式学术合作与交流平台。为计算机视觉在中国的发展与人才的启蒙、培养做出了贡献。研究院的一个先期项目是收集大量的图像，手工

标注图像中的场景、物体和部件、关系、功能等。其标注的广度和精细程度为世界领先，并推动了计算机视觉的物体识别和图像解译任务的发展。2017年7月，朱松纯在美国洛杉矶创立暗物智能科技，并于2018年7月落户广州南沙。

#### 相关论文精选

1. 论文标题: *Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation*

论文作者: **Zhu Chun Song**, Yuille Alan

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., (1996): 884-900

论文链接: <https://www.aminer.cn/pub/53e9a350b7602d9702c5a6e4/>

2. 论文标题: *Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling*

论文作者: **Zhu Chun Song**, Wu Yingnian, Mumford David

论文出处: International Journal of Computer Vision, (1998): 107-126

论文链接: <https://www.aminer.cn/pub/53e9a79eb7602d97030da851/>

3. 论文标题: *Image segmentation by data driven Markov chain Monte Carlo*

论文作者: Tu Zhuowen, **Zhu Song-Chun**, Shum Heung-Yeung

论文出处: Pattern Analysis and Machine Intelligence, IEEE Transactions, Volume 24, Issue 5, (2002): 657-673.

论文链接: <https://www.aminer.cn/pub/53e9a0d1b7602d97029b95b1/>

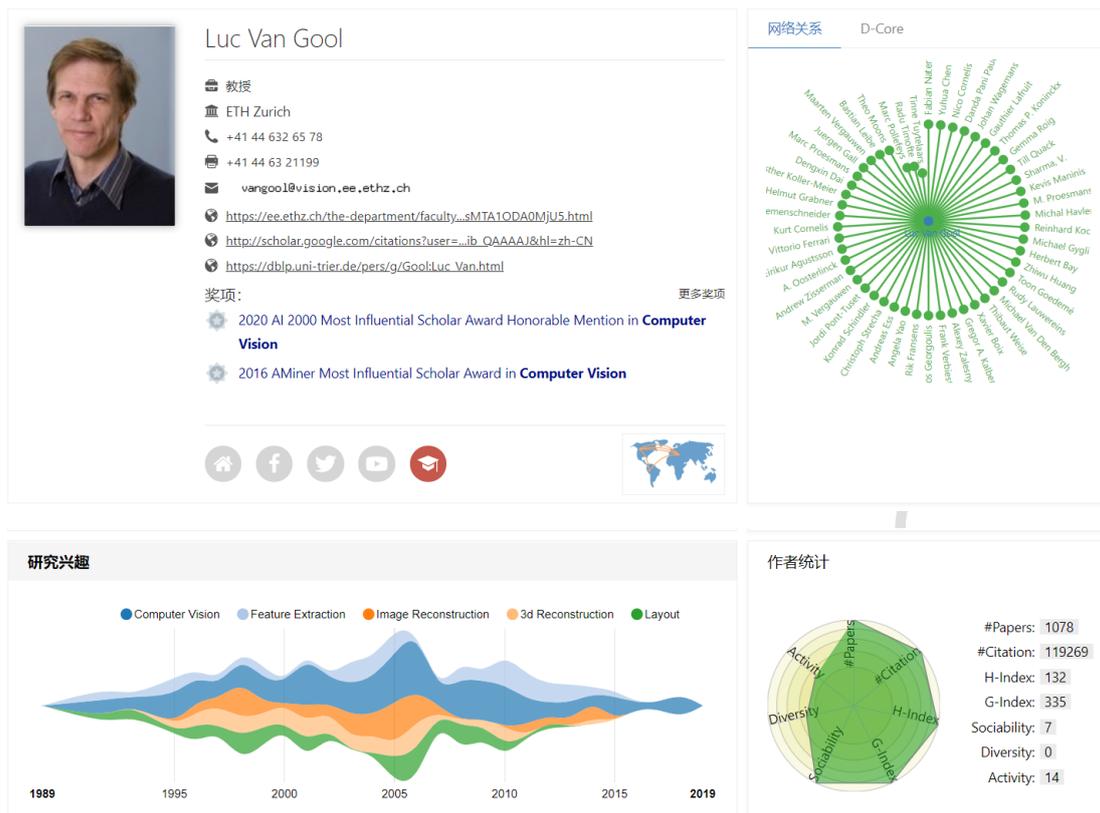
#### ●**Luc Van Gool**

教授

苏黎世联邦理工学院

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Computer Vision

2016 AMiner Most Influential Scholar Award in Computer Vision



Luc Van Gool, 苏黎世理工学院教授, 电气工程系计算机视觉小组组长。

Luc Van Gool 和他的研究团队的项目包括: 欧盟行动计划先锋项目、欧盟 Esprit 项目以及 Soquetec 项目。他致力于从基础研究到应用驱动的开发。他的主要研究方向包括二维和三维物体识别、纹理分析、距离获取、立体视觉、机器人视觉和光流。Luc Van Gool 是一些主要国际会议的委员会成员, 包括 ICCV、ECCV 和 CVPR。

### 相关论文精选

#### 1. 论文标题: *Speeded-Up Robust Features (SURF)*

论文作者: Bay Herbert, Ess Andreas, Tuytelaars Tinne, **Gool Van Luc**

论文出处: *Computer Vision and Image Understanding*, (2008): 346-359

论文链接: <https://www.aminer.cn/pub/53e9b54fb7602d9704084778/>

#### 2. 论文标题: *The Pascal Visual Object Classes (VOC) Challenge*

论文作者: Everingham Mark, **Gool Luc**, Williams K. Christopher, Winn John, Zisserman Andrew

论文出处: *International Journal of Computer Vision*, (2010): 303-338

论文链接: <https://www.aminer.cn/pub/53e99fc2b7602d9702899ee6/>

3. 论文标题: *A comparison of Affine Region Detectors*

论文作者: Mikolajczyk K., Tuytelaars T., Schmid C., Zisserman A., Matas J., Schaffalitzky F., Kadir T., **Gool V. L.**

论文出处: International Journal of Computer Vision, Volume 65, Issue 1, 2004.

论文链接: <https://www.aminer.cn/pub/53e99ca0b7602d9702550b5a/>

● **Takeo Kanade**

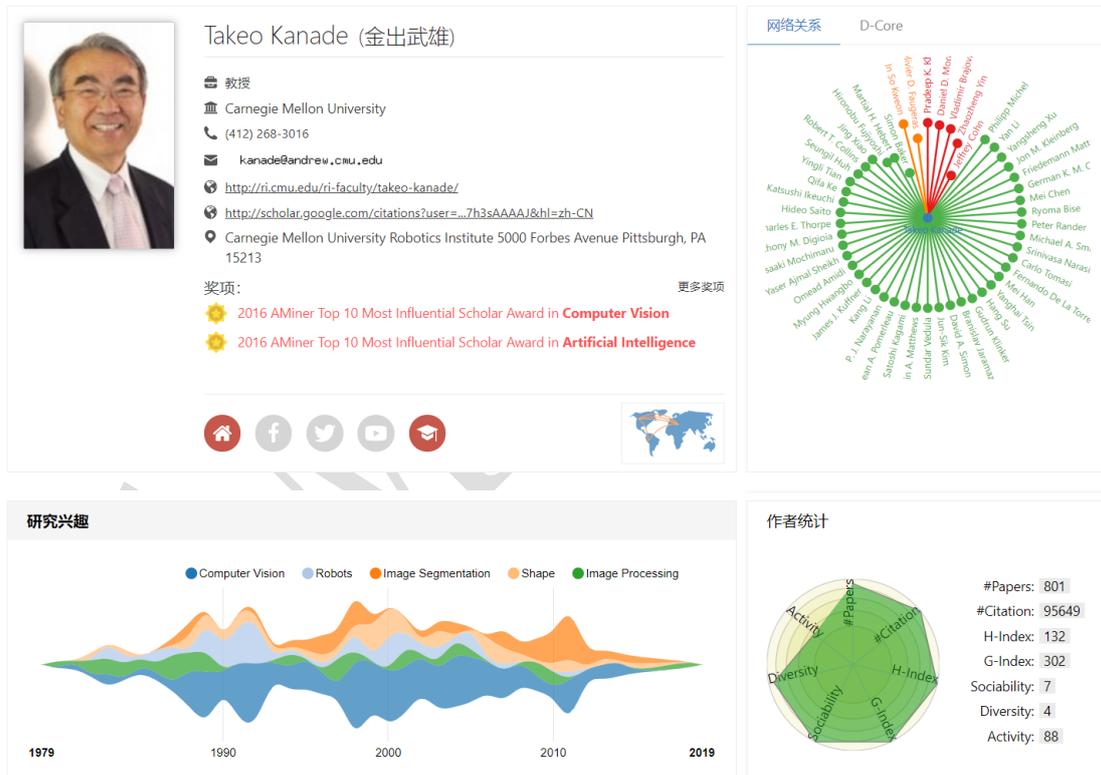
教授

卡内基梅隆大学

2016 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2016 AMiner Top 10 Most Influential Scholar Award in Artificial Intelligence

2018 AMiner Most Influential Scholar Award in Virtual Reality



Takeo Kanade, 卡内基梅隆大学生命质量技术工程研究中心主任。在京都大学信息科学系担任教员后, 他于 1980 年加入卡内基梅隆大学。1992 年至 2001 年任机器人研究所所长, 他建立了东京数字人类研究中心。

Takeo Kanade 的研究兴趣是计算机视觉、视觉和多媒体技术以及机器人技术。Takeo Kanade 在进行研究时强调的主题是: 运用感知和控制过程中涉及到的

物理、几何和语义特性来创建智能机器的声音理论的制定，以及基于这些理论的工作系统的演示。

Takeo Kanade 目前的项目包括计算机视觉（运动、立体和目标识别）、面部表情识别、虚拟现实、基于内容的视频和图像检索、基于 VLSI 的计算传感器、医疗机器人和自主直升机的基础研究和系统开发。

### 相关论文精选

1. 论文标题: *An Iterative Image Registration Technique with an Application to Stereo Vision*

论文作者: Lucas D. Bruce, **Kanade Takeo**

论文出处: IJCAI, (1981): 674-679

论文链接: <https://www.aminer.cn/pub/53e9b9cdb7602d97045c7cb1/>

2. 论文标题: *Shape and Motion from Image Streams under Orthography: A Factorization Method*

论文作者: Tomasi Carlo, **Kanade Takeo**

论文出处: International Journal of Computer Vision, (1992): 137-154

论文链接: <https://www.aminer.cn/pub/53e9b6fab7602d97042935b8/>

3. 论文标题: *Comprehensive Database for Facial Expression Analysis*

论文作者: **Kanade Takeo**, Tian Yingli, Cohn F. Jeffrey

论文出处: FG, (2000): 46-46

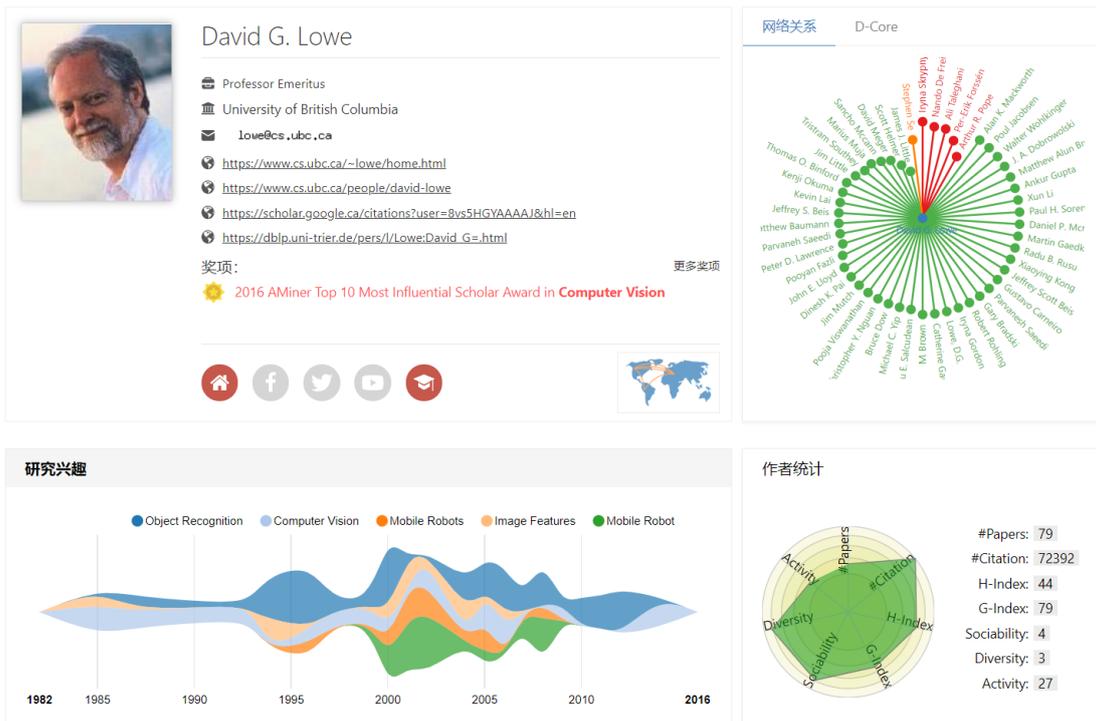
论文链接: <https://www.aminer.cn/pub/53e9a3e7b7602d9702d00f5f/>

### ●David G. Lowe

教授

加拿大英属哥伦比亚大学

2016 AMiner Top 10 Most Influential Scholar Award in Computer Vision



David Lowe, 加拿大英属哥伦比亚大学计算机科学系名誉教授。从 2015 年到 2018 年, David Lowe 是谷歌 Machine Intelligence Group 的资深研究科学家。2009 年至 2015 年, 他是 Cloudburst Research 的联合创始人和董事长, Cloudburst Research 是一家位于温哥华的计算机视觉初创公司, 于 2015 年 5 月被谷歌收购。1987 年至 2015 年, 他是不列颠哥伦比亚大学计算机科学教授。1984 年至 1987 年, 他在纽约大学库兰特数学科学研究所任计算机科学助理教授。

在 2011 年和 2017 年的国际计算机视觉会议上, David Lowe 获得了 Helmholtz 奖, 并在 2015 年获得了 PAMI 杰出研究员奖。他的研究兴趣包括计算机视觉、机器学习和大脑的计算模型。

### 相关论文精选

1. 论文标题: *Distinctive Image Features from Scale-Invariant Keypoints*

论文作者: **Lowe G. David**

论文出处: International Journal of Computer Vision, (2004): 91-110

论文链接: <https://www.aminer.cn/pub/53e9986eb7602d97020ab93b/>

2. 论文标题: *Automatic Panoramic Image Stitching using Invariant Features*

论文作者: Brown Matthew, **Lowe G. David**

论文出处: International Journal of Computer Vision, (2007): 59-73

论文链接: <https://www.aminer.cn/pub/53e9a21db7602d9702b23b5c/>

3. 论文标题: *Perceptual Organization and Visual Recognition*

论文作者: **Lowe G. David**

论文出处: *Perceptual Organization and Visual Recognition*, 1985.

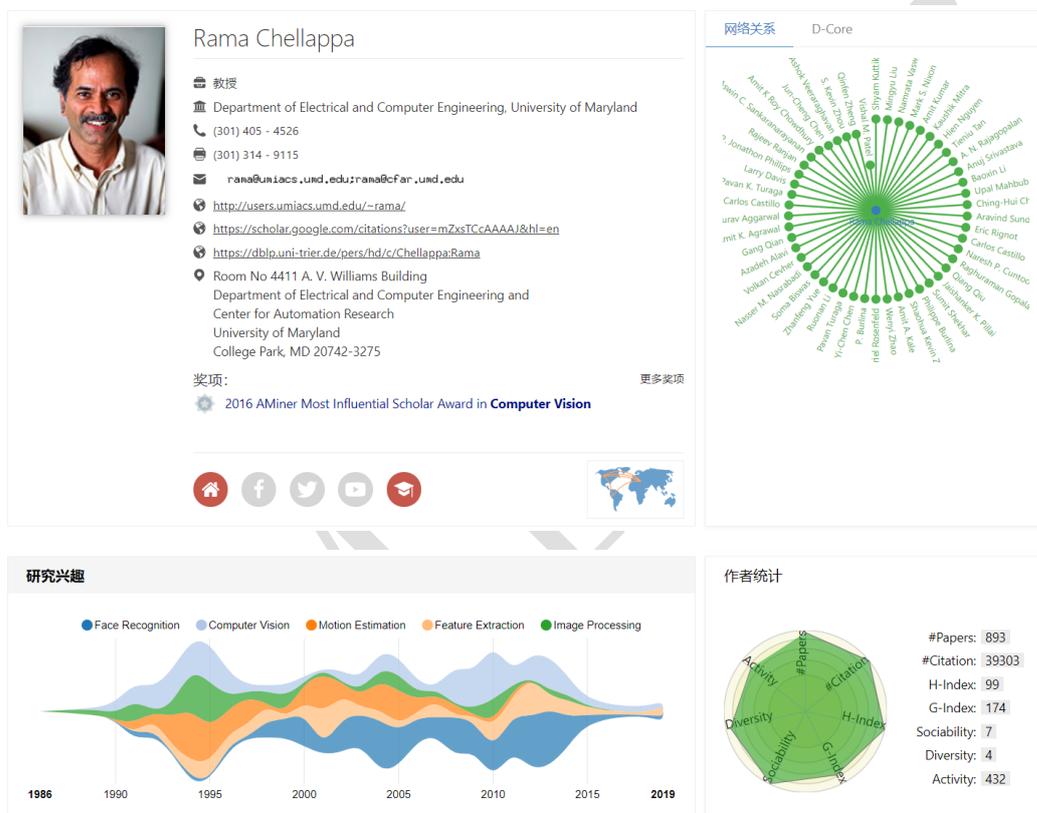
论文链接: <https://www.aminer.cn/pub/53e9aecbb7602d97038f1a5f/>

### ● Rama Chellappa

教授

马里兰大学

2016 AMiner Most Influential Scholar Award in Computer Vision



Rama Chellappa, 美国马里兰大学 (UMD) 的电气和计算机工程教授和计算机科学教授。他还隶属于美国自动化研究中心和高级计算机研究所, 并担任欧洲经委会部门主席。2005 年, Rama Chellappa 被任命为明塔马丁工程教授。他目前的研究兴趣涉及图像处理、计算机视觉和模式识别等多个领域。Rama Chellappa 获得过 NSF 总统青年研究员奖和四个 IBM 教员发展奖, 国际模式识别协会 (IAPR) 颁发的两项论文奖和 K.S.Fu 奖。他是 IEEE 信号处理协会颁发的社会、技术成就和功勋服务奖的获得者, 还获得了 IEEE 计算机协会颁发的技术成就和优秀服务奖。2010 年, 他被普渡大学评为杰出的欧洲经委会成员。Rama Chellappa 曾担任

模式分析和机器智能的主编，以及若干 IEEE 国际和国家会议和研讨会的一般和技术项目主席/联合主席。他是 IEEE 计算机协会的核心成员，曾任 IEEE 信号处理协会杰出讲师和 IEEE 生物测定理事会主席，也是 IEEE、IAPR、OSA、AAAS 和 ACM 的成员。

相关论文精选

1. 论文标题: *Machine Recognition of Human Activities: A Survey*

论文作者: Turaga K. Pavan, **Chellappa Rama**, Subrahmanian S. V., Udrea Octavian

论文出处: IEEE Trans. Circuits Syst. Video Techn., (2008): 1473-1488

论文链接: <https://www.aminer.cn/pub/53e9b2eab7602d9703da3962/>

2. 论文标题: *Discriminant Analysis of Principal Components for Face Recognition*

论文作者: Zhao W, Krishnaswamy A, **Chellappa R**, Swets DL, Weng J

论文出处: IEEE International Conference on Automatic Face and Gesture Recognition, (1998): 336-341,

论文链接: <https://www.aminer.cn/pub/5b076eb4da5629516ce743a6/>

3. 论文标题: *Visual Tracking and Recognition Using Appearance-adaptive Models in Particle Filters*

论文作者: Zhou Kevin Shaohua, **Chellappa Rama**, Moghaddam Baback

论文出处: IEEE Transactions on Image Processing, (2004): 1491-1506

论文链接: <https://www.aminer.cn/pub/53e9bb7ab7602d97047bdd4f/>

●Larry S. Davis

教授

马里兰大学

2016 AMiner Most Influential Scholar Award in Computer Vision

**Larry S. Davis**

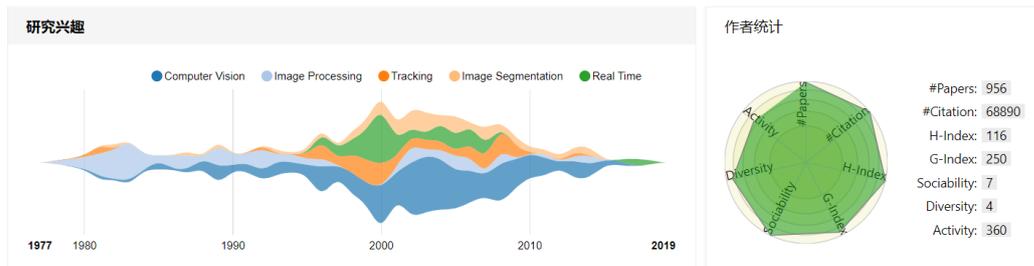
教授

Department of Computer Science, University of Maryland  
 (301) 405-6718;(301)405-2662  
 lsd@cs.umd.edu; lsdavis@umd.edu  
<http://users.umiaccs.umd.edu/~lsd/>  
<https://www.cs.umd.edu/people/lisdavis>  
<https://scholar.google.com/citations?user=ARagAAAA&hl=zh-CN>  
 IRB 4248

奖项:

- 2016 AMiner Most Influential Scholar Award in Computer Vision

网络关系 D-Core



Larry S. Davis, 美国计算机科学家、马里兰大学帕克分校教授。1977 年至 1981 年，他是德克萨斯大学奥斯汀分校计算机科学系的助理教授。1981 年，他在马里兰大学担任教授。1985 年至 1994 年任马里兰大学高级计算机研究所所长。目前 Larry S. Davis 是计算机科学系和高级计算机研究所的教授。Larry S. Davis 最出名的是他在计算机视觉领域的研究。他是国际模式识别协会（IAPR）、电气与电子工程师协会（IEEE）和计算机协会（ACM）的成员。

### 相关论文精选

1. 论文标题: *W/sup 4/: Real-time Surveillance of People and Their Activities*

论文作者: Haritaoglu Ismail, Harwood W J D, **Davis S Larry**

论文出处: IEEE Transactions on Pattern Analysis and Machine Intelligence, (2000): 809-830

论文链接: <https://www.aminer.cn/pub/56d835a9dabfae2eee3def01/>

2. 论文标题: *W4: Real-Time Surveillance of People and Their Activities*

论文作者: Aritaoglu Ismail, Harwood Davis, **David S. Larry**

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., (2000): 809-830

论文链接: <https://www.aminer.cn/pub/53e9ad92b7602d970378a6d0/>

3. 论文标题: *Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance*

论文作者: A. Elgammal, R. Duraiswami, D. Harwood, **L.S. Davis**

论文出处: Proceedings of the IEEE, Volume 90, Issue 7, 2002: 1151-1163.

论文链接: <https://www.aminer.cn/pub/5582c82c0cf2a81de2664fa7/>

### ●Antonio Torralba

教授

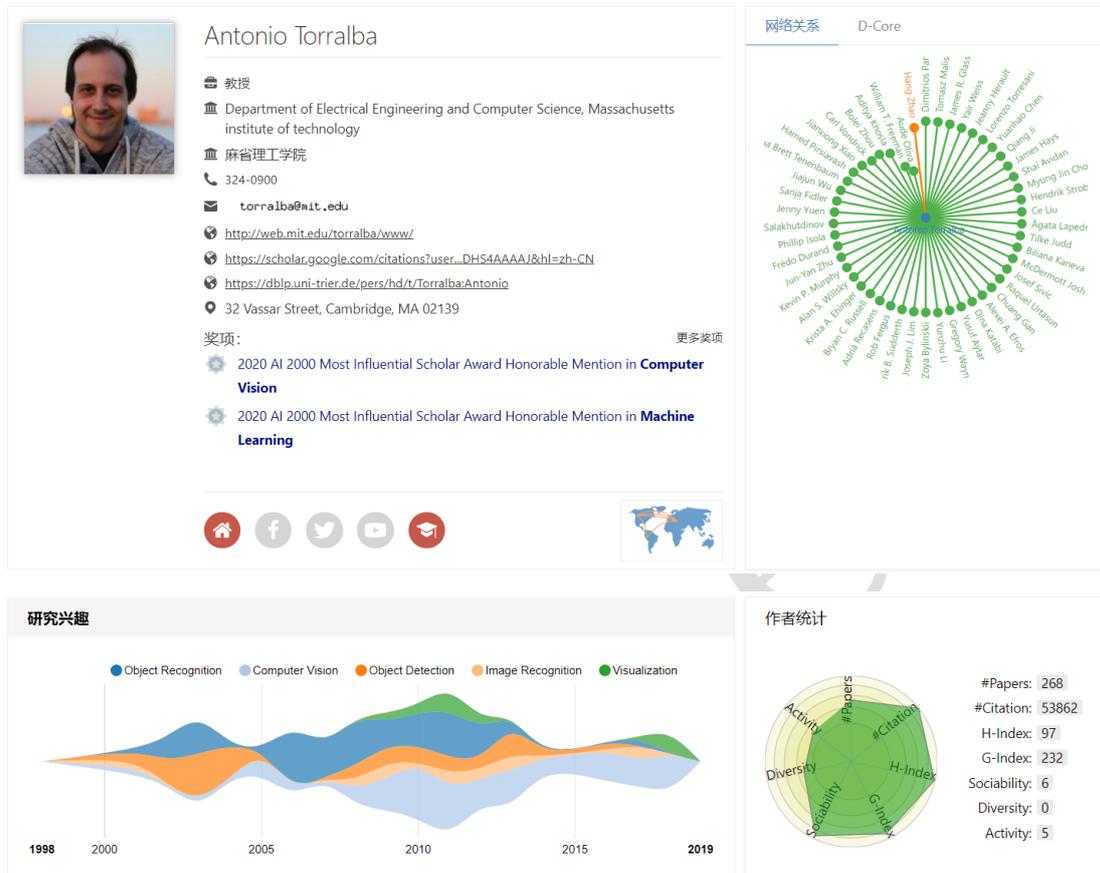
麻省理工学院

2018 AMiner Top 10 Most Influential Scholar Award in Computer Vision

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Computer Vision

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Machine Learning

2018 AMiner Most Influential Scholar Award in Machine Learning



Antonio Torralba, 1994 年获得西班牙电信 BCN 电信工程学位, 2000 年获得法国格勒诺布尔国家理工学院信号、图像和语音处理博士学位, 是计算机视觉, 机器学习和人类视觉感知方面的专家, 是麻省理工电气工程和计算机科学系的教授, 也是 MIT-IBM Watson AI Lab 的 MIT 主任。

Antonio Torralba 是“计算机视觉国际期刊”的副主编, 并在 2015 年担任计算机视觉和模式识别会议的计划主席。他获得了 2008 年国家自然科学基金会职业奖, 这是 IEEE 计算机视觉会议的最佳学生论文奖。2017 年, 他获得了 Frank Quick Faculty 研究创新奖学金和 Louis D. Smullin('39)卓越教学奖。

### 相关论文精选

1. 论文标题: *Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope*

论文作者: Oliva Aude, **Torralba Antonio**

论文出处: International Journal of Computer Vision, (2001): 145-175

论文链接: <https://www.aminer.cn/pub/53e9b1eab7602d9703c7f96b/>

2. 论文标题: *Learning Deep Features for Scene Recognition using Places Database*

论文作者: Zhou Bolei, Garcia Lapedriza Agata, Xiao Jianxiong, **Torralla Antonio**,  
Oliva Aude

论文出处: NIPS, (2014): 487-495

论文链接: <https://www.aminer.cn/pub/555048ef45ce0a409eb72b8f/>

3. 论文标题: *Learning to Predict Where Humans Look*

论文作者: Judd Tilke, Ehinger A. Krista, Durand Frédo, **Torralla Antonio**

论文出处: ICCV, (2009), 2106-2113

论文链接: <https://www.aminer.cn/pub/53e99db7b7602d9702674ca0/>

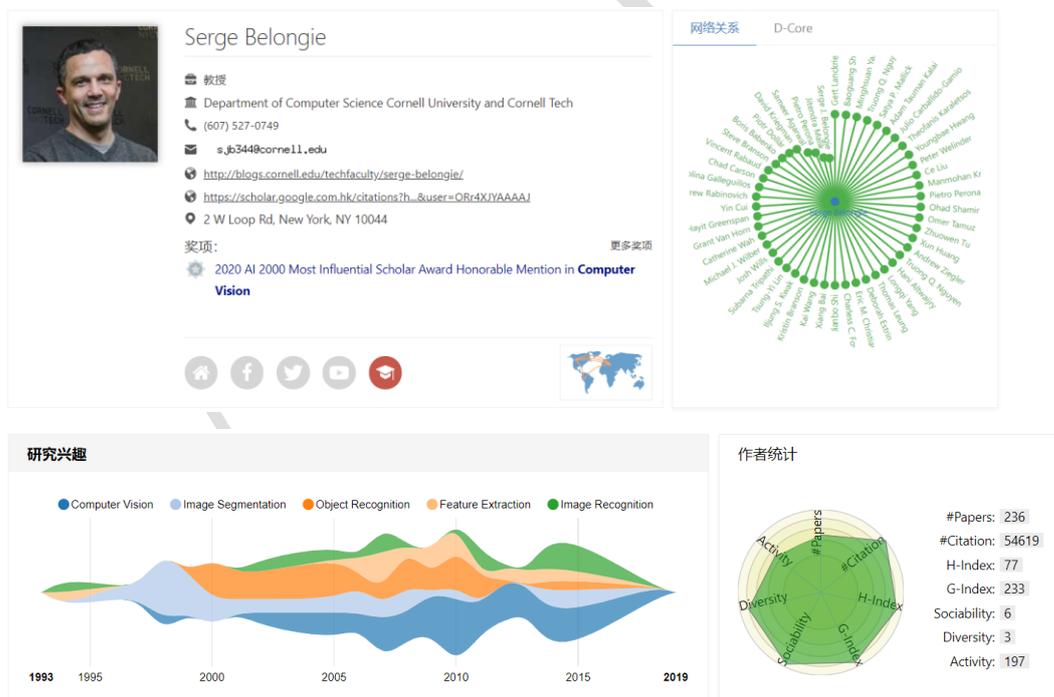
### ●Serge Belongie

教授

康奈尔大学

2020 AI 2000 Most Influential Scholar Award Honorable Mention in Computer Vision

2018 AMiner Most Influential Scholar Award in Computer Vision



Serge Belongie, 康奈尔大学计算机科学系的教授, 也是康奈尔科技公司副院长。他于 1995 年获得加州理工学院电子工程学士学位, 并于 2000 年获得伯克利大学电子工程博士学位。在伯克利期间, 他的研究得到了 NSF 研究生奖学金的

支持。2001 年至 2013 年，他是圣地亚哥加利福尼亚大学计算机科学与工程系的教授。

他的研究兴趣包括计算机视觉、机器学习。他还是数家公司的联合创始人，其中包括 Digital Persona、Anchovi Labs（由 Dropbox 收购）和 Orpixon。他是 NSF 职业奖、Alfred P.Sloan 研究奖学金和麻省理工学院技术评论“35 岁以下创新者”奖的获得者。

### 相关论文精选

1. 论文标题: *Microsoft COCO: Common Objects in Context*

论文作者: Lin Tsung-Yi, Maire Michael, **Belongie Serge**, Hays James, Perona Pietro, Ramanan Deva, Dollár Piotr, Zitnick Lawrence C

论文出处: ECCV (5), (2014): 740-755

论文链接: <https://www.aminer.cn/pub/53e9a479b7602d9702d98afa/>

2. 论文标题: Shape Matching and Object Recognition Using Shape Contexts

论文作者: **Belongie S.**, Malik J., Puzicha J.

论文出处: IEEE Trans. Pattern Anal. Mach. Intell., (2002), 509-522

论文链接: <https://www.aminer.cn/pub/53e99d74b7602d9702630723/>

3. 论文标题: *Feature Pyramid Networks for Object Detection*

论文作者: Lin Tsung-Yi, Dollár Piotr, Girshick B. Ross, He Kaiming, Hariharan Bharath, **Belongie J. Serge**

论文出处: CVPR, 2017.

论文链接: <https://www.aminer.cn/pub/58d82fc8d649053542fd5862/>

## 3.3 计算机视觉代表性团队介绍

### ●微软亚洲研究院视觉计算组

多年来，微软亚洲研究院在计算机视觉领域长期耕耘，做出了多项重大科研贡献，包括用于深度神经网络的 ResNet（残差网络），用于实时物体检测的 Faster R-CNN，用于立体视觉的置信度传播算法，用于图像分割的 Lazy napping 算法以及暗通道去雾法等等。特别是在 2015 年的 ImageNet 计算机视觉识别挑战赛中，微软亚洲研究院的技术率先超越了人类的图像识别水平，取得了里程碑式的突破。ResNet 已经成为计算机视觉乃至人工智能领域的一大经典技术，为业界广泛使用。不仅如此，微软亚洲研究院的计算机视觉技术对微软核心产

品也有颇多贡献，例如 Windows 中的人脸识别，OneNote 中的文字识别，以及微软必应搜索和微软认知服务中的图像识别技术等等。

#### 代表人物：

**郭百宁**，现为微软亚洲研究院常务副院长，负责图形图像领域的研究工作。他于 1999 年加盟微软中国研究院（微软亚洲研究院前身），此前是美国英特尔公司硅谷总部研究院的资深研究员。他还是电气电子工程师学会会士（IEEE Fellow）和美国计算机协会会士（ACM Fellow）。郭百宁士的研究兴趣包括计算机图形学、计算机可视化、自然用户界面以及统计学习。他在纹理映射建模、实时渲染以及几何模型等领域取得的研究成果尤为突出。

**王井东**，2001 和 2004 年在清华大学自动化系先后获得学士学位和硕士学位，2007 年在香港科技大学计算科学与工程系获得博士学位。2007 年进入微软亚洲研究院，现任视觉计算组资深研究员。研究兴趣包括计算机视觉、多媒体以及机器学习。目前他研究的问题包括高效神经网络结构的设计、行人再识别以及多媒体搜索等。

**团队网站：**<https://www.microsoft.com/en-us/research/group/visual-computing/>

#### ●中国科学院 自动化研究所 智能感知与计算研究中心

智能感知与计算研究中心为中科院自动化研究所独立建制的科研部门，致力于研究泛在智能感知理论与技术以及与之相伴的海量感知数据的智能分析与处理。中心目前主要在多模态智能计算、生物识别与安全、生物启发的智能计算、智能感知基础理论四个方面展开科学研究。

智能感知与计算中心在相关研究领域已累计发表 900 余篇期刊与会议文章，已授权及公开专利 200 余项，曾经获得国家技术发明二等奖、国家科技进步二等奖、国家自然科学基金二等奖、北京市科学技术一等奖、全国发明专利优秀奖等。同时，中心在国际前沿的多次竞赛中屡创佳绩，多次夺魁，曾连续两年获得 PASCAL VOC 目标检测竞赛的冠军，蝉联 NICE 系列国际虹膜识别算法竞赛的冠军。

**谭铁牛**，研究员，中国科学院院士、英国皇家工程院外籍院士、发展中国家科学院（TWAS）院士，巴西科学院通讯院士。现为中央人民政府驻香港特别行政区联络办公室副主任、智能感知与计算研究中心主任、中国图象图形学学会理事长。谭铁牛博士主要从事图像处理、计算机视觉和模式识别等相关领域的研究

工作，目前的研究主要集中在生物特征识别、图像视频理解和信息内容安全等三个方向。他主持过一批由国家基金委、国家杰出青年基金、国家 973 计划、863 计划、国际合作计划等资助的科研项目。现已出版编（专）著 15 部并在主要的国内外学术期刊和国际学术会议上发表论文 600 多篇，获准和申请发明专利 140 多项。还曾获中国青年五四奖章、中国青年科技奖以及国家自然科学基金二等奖、国家技术发明二等奖和国家科技进步二等奖各 1 项。

**王亮**，研究员，国际模式识别学会会士，国际电子电气工程师学会会士。目前是模式识别国家重点实验室副主任，中科院自动化所智能感知与计算研究中心常务副主任，中国计算机学会计算机视觉专委会秘书长，中国图像视频大数据产业技术创新战略联盟秘书长，中国图象图形学学会视觉大数据专委会主任，中科院脑科学与智能技术卓越中心骨干人才，中国科学院大学特聘岗位教授，国家重点研发计划云计算与大数据重点专项项目首席科学家，中科人工智能创新技术研究院院长等。王亮主要从事计算机视觉、模式识别、机器学习、数据挖掘等相关领域的研究，承担国家级、省部级各类科研项目 20 余项。已与华为、爱奇艺、腾讯、京东金融、高德、美的等知名企业进行技术研发或应用的广泛合作。部分研究成果已孵化银河水滴科技（北京）有限责任公司。

### ●中国科学院 计算技术研究所 视觉信息处理与学习研究组

视觉信息处理与学习研究组（Visual Information Processing and Learning, VIPL）始建于 1997 年，由高文院士组建。现隶属于中国科学院计算所及中科院智能信息处理重点实验室，同时也是先进人机通信技术联合实验室（JDL）的一部分。目前，VIPL 研究组主要成员包括研究人员 20 余名、博士/硕士研究生 50 余名。研究组的研究领域涵盖计算机视觉、模式识别、机器学习/深度学习、人机交互和情感计算等学科方向，研究组在相关领域发表研究论文 500 余篇，其中 100 余篇发表在计算机学会认定的领域 A 类国际期刊或会议上。研究成果获 2015 年度国家自然科学基金二等奖，2005 年度国家科技进步二等奖等。研究组不仅致力于开展具有高学术影响力的基础理论和方法研究，同时密切关注具有潜在应用价值的技术创新，研发的多项创新技术已获得实际应用。

**代表人物：**

**陈熙霖**，研究员，ACM Fellow, IEEE Fellow, IAPR Fellow，中国计算机学会会士，国家杰出青年基金获得者。主要研究领域为计算机视觉、模式识别、多媒体技术以及多模式人机接口。先后主持多项自然科学基金重大、重点项目、973计划课题等项目的研究。曾任 IEEE Trans. on Image Processing 的 Associate Editor，目前是 IEEE Trans. on Multimedia 和 Journal of Visual Communication and Image Representation 的 Associate Editor、Journal of Computer Science and Technology 领域编委、计算机学报副主编、人工智能与模式识别副主编，担任过 FG2013 / FG2018 General Chair 以及 CVPR 2017 / 2019 / 2020，ICCV 2019 等的 Area Chair。

**山世光**，研究员，中科院智能信息处理重点实验室常务副主任。他的研究领域为计算机视觉和机器学习。已在国内外刊物和学术会议上发表论文 300 余篇，其中 CCF A 类论文 80 余篇，论文被谷歌学术引用 18000 余次。所研发的人脸识别相关研究成果获 2005 年度国家科技进步二等奖（第 3 完成人），在高维、非线性视觉模式分析方面的研究成果获 2015 年度国家自然科学基金二等奖（第 2 完成人），视觉流形建模与学习方面的研究成果获 CVPR2008 Best Student Poster Award Runner-up 奖。他带领团队研发的人脸识别技术已应用于公安部门、华为等众多产品或系统中，取得了良好的经济和社会效益。曾应邀担任过 ICCV11，ACCV12/16/18，ICPR12/14/20，FG13/18/20，ICASSP14，BTAS18，CVPR19/20/21 等十余次领域主流国际会议的领域主席。他是基金委优青，国家重要人才计划入选者，科技部创新人才推进计划中青年科技创新领军人才，人社部国家百千万人才工程有突出贡献中青年专家，CCF 青年科学家奖获得者，北京市科技新星，中科院青促会优秀会员。

**团队网站：** <http://vipl.ict.ac.cn/>

#### ●北京大学机器感知与智能教育部重点实验室

北京大学机器感知与智能教育部重点实验室以实现高度智能化的机器感知系统为目标，紧密结合国民经济和社会发展的要求，开展机器感知、智能信息处理与认知科学方面的基础与应用基础研究。目前实验室的主要研究方向有感知机理，计算智能与知识发现，视感知，听感知，数字媒体技术，可视化与视觉计算。主要研究内容包括：探索人的感知机理以及感知与智能的关系；研究机器感知与智能的实现方法，扩展人的感知能力；开发应用于国家建设和改善民生的智能感

知系统。实验室的建设目标是力争通过这些方向的前沿性研究工作，将实验室建设成为多学科交叉的、开放的、具有国际竞争力的机器感知与智能研究基地。

### 代表人物：

**查红彬**，长期从事智能科学技术、计算机视觉的研究。在三维视觉计算、环境几何建模、三维物体识别、三维人脸动画和动态目标跟踪等方面取得了进展。先后提出了基于模型的物体识别、三维建模与形状分析、三维动态目标跟踪等新方法，在将图像处理与分析理论应用到智能机器感知系统方面取得了一系列成果。同时，还在高维数据的表达与特征抽取、物体匹配的距离学习等方面有了一定的突破。另外，还将部分研究成果应用到公安刑侦、医疗与文化遗产的应用中，特别是主持的数字龙门石窟项目得到了国内外同行的高度重视，被中央电视台等多种主流媒体宣传。

**林宙辰**，北京大学信息科学技术学院机器感知与智能教育部重点实验室教授。研究领域包括机器学习、模式识别、计算机视觉、图像处理和数值优化。曾经当选 CVPR、ICCV、NIPS 和 AAA 的领域主席，AAAI 和 IJCAI 的高级程序委员。他还是 IEEE Transactions on Pattern Analysis and Machine Intelligence 和 International Journal of Computer Vision 的编委，IAPR 和 IEEE 会士。

**团队网站：**<http://robotics.pkusz.edu.cn/>

### ●清华大学 自动化系 信息处理研究所

信息处理研究所拥有“模式识别与智能系统”和“生物信息学”两个学科专业。现有教授 8 人，副教授 2 人，讲师 4 人，工程师 1 人。其中中科院院士一名（李衍达教授），博士生导师 8 人，教育部长江学者特聘教授 1 人，国家杰出青年基金获得者 2 人，新世纪优秀人才支持计划获得者 3 人，全国百篇优秀博士学位论文获得者 3 人，另有教育部长江讲座教授 1 人。经过多年的发展，研究所形成了以智能信息处理理论、方法与应用为核心的 3 个主要研究方向：模式识别与机器学习、生物信息学与计算生物学、智能信号处理理论及应用，同时积极探索新兴的现代服务科学研究方向。研究内容包括模式识别、机器学习、图像分析、信号处理的理论、方法与技术，智能信息处理在机器视觉、智能交通系统、石油物探、雷达、数字通讯、现代服务业等多个领域中的应用，以及基于信息理论与技术的生命科学基本问题的探索。这些方向之间既有各自鲜明的特色，又有紧密的内在联系，相互促进、相互渗透，形成了一个有特色、高水平的模式识别与智

能系统学科。近 5 年来，承担 973、863 和自然科学基金项目 30 多项，获得国家和部委级科研和教学奖励 10 多项，在国际期刊上发表高水平论文 110 多篇，在全国重点学科评估中名列全国第一。

**周杰**，清华大学自动化系主任，教授，博士生导师，国家杰出青年基金获得者。现担任中国自动化学会副秘书长兼教育工作委员会主任、中国人工智能学会模式识别专业委员会副主任、模式识别国家重点实验室学术委员会委员等职务，是多家国际国内学术刊物的编委。研究方向包括模式识别、计算机视觉、数据分析与挖掘等，在国内外学术刊物及国际会议上发表论文 150 多篇，其中 IEEE 期刊 30 多篇（其中 IEEE T-PAMI 长文 10 篇）；已获授权美国发明专利 6 项、中国发明专利 16 项。

**鲁继文**，清华大学自动化系副教授、博士生导师，中组部青年千人计划入选者，国家优秀青年基金获得者。主要研究方向为计算机视觉、机器学习和智能机器人，发表 IEEE 汇刊论文 60 余篇（其中 PAMI 论文 11 篇），谷歌学术引用 6100 余次（H 指数为 40），获 IEEE 国际会议最佳论文奖 2 次。主持国家重点研发计划课题 1 项、国家自然科学基金项目 2 项。曾/现任 IEEE Transactions on Circuits and Systems for Video Technology、IEEE Transactions on Biometrics, Behavior, and Identity Science、Pattern Recognition 等 7 个国际期刊编委，中国工程院院刊 Engineering 青年通讯专家，ACCV、ICIP、ICME 等 20 多个国际会议的领域主席。担任 IEEE 信号处理学会多媒体信号处理技术委员会委员，IEEE 信号处理学会信息取证与安全技术委员会委员，IEEE 电路与系统学会多媒体系统与应用技术委员会委员。

**团队网站：**<https://www.au.tsinghua.edu.cn/publish/au/1089/index.html>



# 4. 应用篇





计算机视觉领域内的大部分企业都提供 API、SDK、定制化解决方案的产品交付形式，但对行业用户来说，API 与 SDK 产品需要进行二次开发与调整使用难度较大，效果与体验不好，因此行业用户普遍青睐定制解决方案。本篇将从定制解决方案角度来描述计算机视觉技术的主要应用领域，具体包括：城市公共安全、政务民生、金融服务、新零售等，并在章节的最后畅想了计算机视觉应用的未来发展。

## 4.1 城市公共安全

伴随着人工智能软硬件技术的不断发展，计算机视觉技术早已突破了肉眼精度的图像识别，并广泛应用于城市公共安全领域，包括：车辆识别、安防监控、刑侦追逃、罪犯识别、边防安全检查、金融安保、交通监管、监狱人员管理等诸多场景，保卫着城市的公共安全。

城市公共安全是指一种安全概念，即将来自各种传感器、输入端和数据源的关键物理安全信息集成综合到一个 IT 平台上，使城市能够提前预测威胁并做出及时响应和合理应对。城市公共安全解决方案的核心是视频监控系统、应急指挥中心及 LTE（长期演进）等关键通信技术。一般来说，视频监控系统为系统提供主要输入信息；应急指挥中心提供分析和应对各种威胁的手段；关键通信系统则确保在现场以及应急指挥中心的相关人员保持良好的通信网络。

### 应用实例

依图科技公司开发的视频图像车辆识别系统就已多次帮助警方破获案件，当有嫌犯驾车作案，警方可以运用“车辆识别系统”进行品牌过滤，快速识别出车辆，短时间内神速破案。与车辆等物体识别相比，人脸识别的应用范和场景更加广泛，当前技术已经可以做到在上亿数量级的人像库中进行人脸静态比对识别，协助警务人员快速定位嫌犯过往在摄像头下出现的时间和地点轨迹，继而控制嫌犯。

计算机视觉技术在公共安防领域的快速进展得益于两方面因素支持。一是在技术层面，视频监控的 IP 化、高清化使得视频数据规模和可获得性得到显著改善，同时深度学习算法和并行计算芯片的出现，消除了智能视频分析的部分技术障碍。二是在市场层面，我国多部委陆续推出“平安城市”“天网工程”“雪亮

工程”等重大系统性安防工程，推动了视频监控基础设施的快速建设和完善，有力推动了计算机视觉技术在公共安全领域的应用。

### 4.2 政务民生

目前，计算机视觉技术在政务民生场景的应用已出现在审批、核准、备案、证明等多项行政权力办理事件中。企业和个人在线申请，政府在线办理，全面实现网上政务公开和互动交流等便民应用，同时基于大数据的人工智能，助力政府制定更加精准、灵活、前置性的决策方案。

#### 应用实例

旷视科技可以利用人脸识别技术为公安部提供“网上身份证”支持。有了网上身份证，每个人都可以在网上生成一本终身唯一编号的“身份证网上副本”，今后办理一些实名认证业务时即可“刷脸”完成认证，不用再携带实体身份证。通过多因子认证技术实现互联网上的“实名+实人+实证”的真实身份认证，在保护公民隐私信息的同时有效解决了“我就是我”的问题，让市民在网上办事变得更加可靠、安全。

其他具有代表性的智能政府还包括深圳公安局将传统的窗口“面对面”排队向网上办理转变，“刷脸”就可以进行户政办理，同时基本建成全市统一的政务信息资源共享体系，汇集 29 家单位的 385 类信息资源、38 亿多条数据 27，为政务服务全面智能化提供数据支持。杭州构建一体化的智能电子政务管理体系，数字城管、规划系统、财政系统业务系统在电子政务外网得到整合，并提供一站式服务等。

### 4.3 金融服务

目前，金融服务与计算机视觉等其他人工智能技术深度融合已发成为金融服务行业的常规业态。计算机视觉在金融行业中的应用场景有智能获客，依托大数据，对金融用户进行画像，通过需求响应模型，极大地提升获客效率；身份识别，以人工智能为内核，通过人脸识别、声纹识别、指静脉识别等生物识别手段，再加上各类票据、身份证、银行卡等证件票据的 OCR 识别等技术手段，对用户身份进行验证，大幅降低核验成本，有助于提高安全性；智能客服，基于计算机视觉技术，拓展客服领域的深度和广度，大幅降低服务成本，提升服务体验。

#### 应用实例

人脸识别技术在我国银行业已经完全普及，包括农行、建行、中行、交行等全国 50 多家银行已经利用该技术实现了用户的金融身份认证与远程平台认证。中国农业银行的超级柜台、刷脸取款更是全国首先应用人脸识别技术的四大行之一。由于人脸登录支付功能的支持，人们不再需要输入繁琐的密码，只需对着手机镜头眨眨眼、转转头便可轻松完成登录。

## 4.4 新零售

计算机视觉新零售是以消费者体验为中心，由数据驱动的泛零售形态。新零售的想象空间比较难以把握，因为它的应用场景非常多，涵盖了百货公司、购物中心、便利店，甚至直播、视频、电子商务等等。但无论在哪个场景，买东西的本质都是人和商品。新零售的核心就是理解人，理解物；把人和物联系起来，让用户更好更快地找到他满意的商品，或者更好更快地把商品送到用户手里。

### 应用实例

计算机视觉技术在新零售领域主要有三方面的应用。

一是用于面部识别。主要是识别用户性别、年龄范围、浏览记录、浏览习惯等基本个人信息，提醒服务提供商和零售商注意及时反应，达到识别老顾客、收获忠诚度、提升广告投放回报率的效果。美食糖果零售商 Lolli&Pops 依靠计算机视觉技术，在老顾客进店的时候就能认出他们，提取其消费记录和个人喜好，提供个性化的商品推荐，顾客可由此获得个性化的消费体验。

二是实现“货架数字化”，推动货架生态变革。零售企业通过计算机视觉技术，实时监控货架上发生的情况，并对缺货、调货等做出反应。亚马逊就会利用计算机视觉技术，依托 RFID、NFC 和各类传感器，跟踪货品，精准管理库存，提升库存管理准确性，节约成本和减少损耗。

三是实现个性化、定制化的营销互动。利用人脸识别、手势识别、图片处理等技术，实现与用户在多种营销场景下的互动，让营销更趋人性化，使得视觉成为品牌连接用户的新流量入口。例如，计算机视觉技术在 Amazon Go 无人商店中就得到了广泛应用，通过智能终端识别物体，结账时不必依靠收银员和条码扫描仪结算。RetailNext 公司的 Aurora 利用计算机视觉技术进行用户行迹策略，捕捉用户脚步、步态、购物线路，促进客户和销售人员之间的互动，提供门店实时服务情况的可视化信息，实现个性化的营销和宣传活动。

## 4.5 产业应用的未来

随着计算机视觉技术与实际应用问题结合的不断深入,通用化的产品越来越难满足不同用户、不同场景的使用需求,主要的行业用户也不具备较强的二次开发能力,因此,定制化的解决方案将成为众多行业用户的主要选择。在教育领域,一些省市已利用人脸识别技术进行考生报名与信息核对;在零售领域无人超市、无人便利店也正在试点运营;此外,3D 视觉、动态视频等新模式也正在被探索发展。

AMiner

# 5. 趋势篇





领域技术分析系统 (<https://analysis.aminer.cn/>) 可以基于 AMiner 的 3 亿篇论文进行深度挖掘, 本篇基于系统全面分析了计算机视觉领域的技术趋势、国际趋势、机构趋势等。

## 5.1 技术趋势

如图 19 所示, 计算机视觉领域的技术趋势具有以下几个特点。首先, 近十年以来卷积神经网络技术在该领域得到快速发展, 在几年时间内火速上升至第一位。大数据时代的到来, 使深度卷积神经网络具备更加丰富的网络结构, 与传统的机器学习相比, 在特征表达与特征学习方面更具优势以深度学习算法深度卷积神经网络模型为基础所提出的计算机视觉领域在识别能力上取得了显著成绩。ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 是近年来计算机视觉领域最受追捧也是最具权威的学术竞赛之一。在 2010 年和 2011 年, ILSVRC 的图像分类错误率徘徊在 26% 左右。自从 2012 年, 来自多伦多大学的一个团队进入了一个卷积神经网络模型 (AlexNet) 进入竞争, 在此之后获胜者一直是卷积神经网络。从目的上分类, 图像分类问题的关注度从 90 年代以来一直保持着上升趋势, 特征提取问题虽近些年来热度被卷积神经网络超过, 但也发展势头良好。

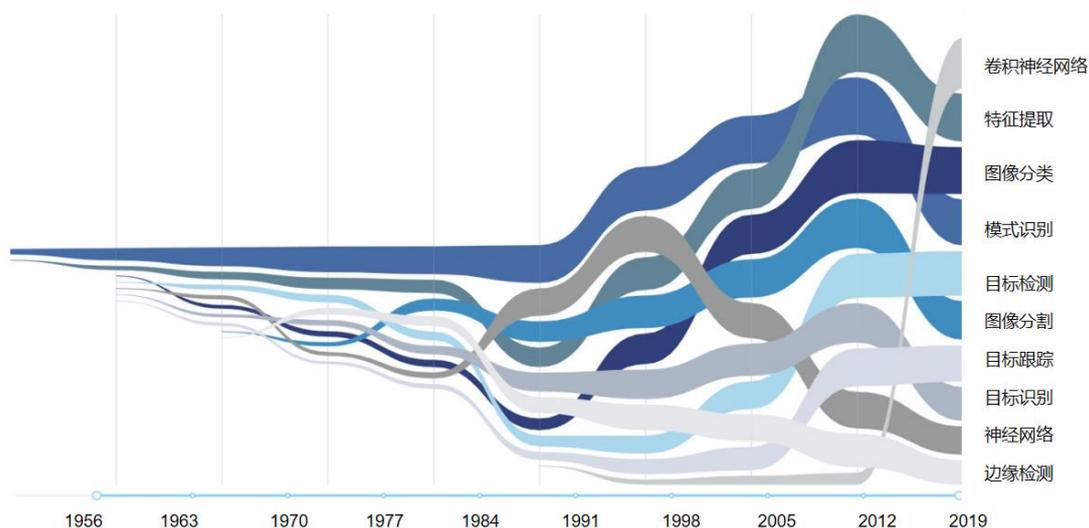


图 19 计算机视觉领域技术趋势图

## 5.2 国家趋势

国家趋势分析如图 20 所示。图中每条色带的宽度表示该国家在当年的研究热度, 与当年该国论文数量呈正相关, 每一年份中按照其热度由高到低进行排序。

国家趋势分析显示当前计算机视觉领域研究热度 TOP10 的国家分别是：中国、美国、印度、英国、日本、德国、加拿大、澳大利亚、法国。其中，中国在计算机视觉领域的发展最令人瞩目，当前研究热度位居领域第一。

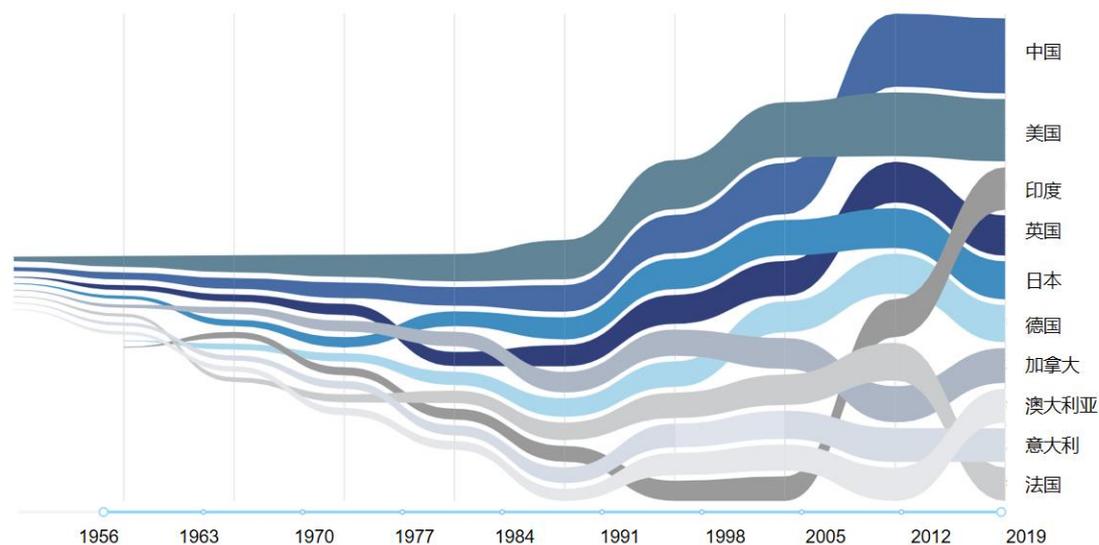


图 20 计算机视觉领域国家趋势图

### 5.3 机构趋势

中国科学院作为国内自然科学最高学术机构，在计算机视觉领域一直处于国内顶尖水平，并且于 2008 年左右在论文数量方面超过加利福尼亚大学伯克利分校，成为世界学术机构发表论文数最多的机构。清华大学同样作为国内最高学府，在计算机视觉领域的发展也令人瞩目，在近几年跃升全球至第二位。如图 21 所示，在排名前十的学术机构中有六所来自中国，包括中国科学院、清华大学、北京大学、浙江大学、哈尔滨工业大学和上海交通大学。相比美国高校在该领域悠久的历史，中国高校在不长的时间内能够后来者居上，可见国家对相关学科领域的支持。前十榜单中除了中国和美国高校，还有新加坡国立大学与南洋理工大学，发展较为稳定。就学术机构排名而言，国家的多样性不大，学术产出依旧集中在少数几个国家中。

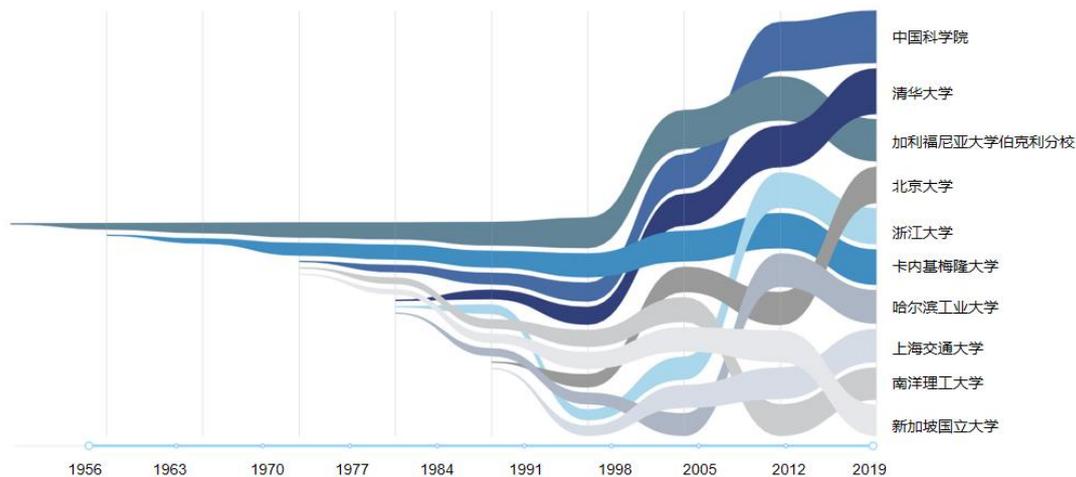


图 21 计算机视觉领域机构趋势图

## 5.4 技术发展面临的挑战

当前，我国计算机视觉技术快速发展，研究热度位居全球首位。技术的发展也带来了就业结构、个人隐私、社会准则等多方面的问题。我国在大力发展计算机视觉技术的同时，技术带来的风险与挑战也同样不容小觑。本小结将从政策标准、隐私安全和技术难点三方面来探讨计算机视觉技术在未来所要面临的挑战。

### （1）政策标准

政府出台的政策标准是保证人工智能应用质量和合法性的社会使能支撑,计算机视觉技术作为人工智能的重要组成,在改变人们的生活方式的同时,大量淘汰传统行业劳动力,带来社会就业结构的不稳定。国家在制定人工智能发展战略的同时,必须加强人工智能标准和规范的制定,出台法律法规并制定伦理道德框架保障人工智能朝着健康的方向发展,开展与人工智能应用相关的民事与刑事责任认定,建立追溯和问责制度,明确人工智能法律主体的权利义务和相应的社会责任,加快研究制定相关安全管理法律法规,出台研发企业与研发人员的道德规则和行为守则,预先对人工智能应用进行收益和风险的评估,研究突发事件的应对措施。

### （2）隐私安全

具有颠覆性和革命性的人工智能技术在侵犯个人隐私、挑战经济安全和社会稳定等方面产生了重要影响。目前针对应用计算机视觉技术的监管制度尚未建立,隐私安全得不到有效保障。在政策标准和监管体系的作用下,建立国家安全和保密领域影响的评估措施,构建安全监测预警机制,出台对数据滥用、侵犯个人隐

私、违背道德伦理等行为的处罚措施。计算机视觉技术涉及的安全隐私风险存在于数据采集、存储和应用等各方面，随着技术的进步，从数据到知识的抽取能力也在快速提升，规范隐私安全保护是人工智能技术面临的主要挑战之一。

### （3）技术难点

从计算机视觉技术的发展角度来看，3D 和 Video 将成为未来主要面对的两大大任务方向，而任务所要面临的挑战主要体现在弱监督、鲁棒性和智能化这三个方面。以下是报告顾问提出的一系列解决办法，以供读者参考。

- Holistic（全盘性）

常见的视觉任务包括分类、物体检测、语义/实例分割、3D 重建等。人类的视觉系统在面向视觉任务时，可以同时全面考虑问题，而非像计算机一样需要拆成若干子任务进行处理。未来对视觉的研究，可模拟人类处理时的方式，将各类子任务融合成一个大任务。

- Cross modality（跨模态学习）

在医学领域，医生做病情诊断时，除了观察 CT、X 光等得到医学影像，还会借助病人病史、化验结果、问询等了解其他信息，这个过程就可以称为跨模态学习。为完成一个最终目标，可能会需要很多其他模态信号加以辅助，视觉任务亦是如此。例如将视频里面的声音与图像进行互监督学习，常见的有篮球、足球等体育比赛中，解说与画面的配合。解说语音提供了很多标签，利用这些标签可帮助训练视觉模型。从大的 AI 角度来说，视觉未来定会和其他模态的数据进行融合。

- X learning（X 学习）

Self-supervised learning（自监督学习）、Unsupervised learning（无监督学习）、Meta learning（元学习）、Life-long learning（终身学习）、Robust learning（稳健学习）、Transfer learning（迁徙学习）等机器学习领域的方法，都有可能在未来视觉研究领域中发挥巨大的作用，催生新的研究方向。

## 参考文献

- [1] 李德毅, 于剑等. 人工智能导论-计算机视觉[M]. 北京: 中国科学技术出版社, 2018.
- [2] Deng J, Dong W, Socher R, et al. Imagenet: A Large-scale Hierarchical Image Database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition IEEE, 2009: 248-255.
- [3] Turk MA, Pentland AP. Face Recognition Using Eigenfaces[C]//Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition IEEE Computer Society, 1991: 586-586.
- [4] Masi I, Wu Y, Hassner T, et al. Deep Face Recognition: A Survey[C]//2018 31st Sibgrapi Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2018: 471-478.
- [5] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks[C]//Neural Information Processing Systems, 2015: 91-99.
- [6] Bergen JR, Anandan P, Hanna KJ, et al. Hierarchical Model-based Motion Estimation[C]//European Conference on Computer Vision, Springer, 1992: 237-252.
- [7] Comaniciu D, Ramesh V, Meer P. Kernel-based Object Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.
- [8] Brox T, Bruhn A, Papenbergh N, et al. High Accuracy Optical Flow Estimation Based on a Theory for Warping[C]//European Conference on Computer Vision, Springer, 2004: 25-36.
- [9] Izadi S, Kim D, Hilliges O, et al. Kinectfusion: Real-time 3d Reconstruction and Interaction Using a Moving Depth Camera[C]//Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, 2011: 559-568.
- [10] Saxena A, Chung SH, Ng AY. 3-d Depth Reconstruction from a Single Still Image[J]. International Journal of Computer Vision, 2008, 76(1): 53-69.
- [11] Park SC, Park MK, Kang MG. Super-resolution Image Reconstruction: A Technical Overview[J]. IEEE Signal Processing Magazine, 2003, 20(3): 21-36.
- [12] Mitchell DP, Netravali AN. Reconstruction Filters in Computer-graphics[J]. ACM Siggraph Computer Graphics, 1988, 22(4): 221-228.

- [13] Schlemper J, Caballero J, Hajnal JV, et al. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction[J]. IEEE Transactions on Medical Imaging, 2017, 37(2): 491-503.
- [14] Vernon D. Machine Vision-automated Visual Inspection and Robot Vision[J]. STIA, 1991, 92(1991): 40499.
- [15] 杨超. 数字图像增强技术研究 with 实现[M]. 电脑编程技巧与维护, 2018, 000(9): 138-139.
- [16] Donoho DL, Flesia AG. Can Recent Innovations in Harmonic Analysis explain key Findings in Natural Image Statistics? [J]. Network: Computation in Neural Systems, 2001, 12(3): 371-393.
- [17] Le Pennec E, Mallat S. Image Compression with Geometrical Wavelets[C]//Proceedings 2000 International Conference on Image Processing (cat. No. 00ch37101), IEEE, 2000: 661-664.
- [18] Le Pennec E, Mallat S. Sparse Geometric Image Representations with Bandelets[J]. IEEE Transactions on Image Processing, 2005, 14(4): 423-438.
- [19] Donoho DL, Huo X. Multiscale and Multiresolution Methods[M], Springer, 2002: 149-196.
- [20] Velisavljevic V, Beferull-lozano B, Vetterli M, et al. Approximation Power of Directionlets[C]//IEEE International Conference on Image Processing 2005, IEEE, 2005: 0-741.
- [21] Krommweh J. Tetrolet Transform: A New Adaptive Haar Wavelet Algorithm for Sparse Image Representation[J]. Journal of Visual Communication and Image Representation, 2010, 21(4): 364-374.
- [22] Candes EJ. Ridgelets: Theory and Applications[D]. Stanford University, 1998.
- [23] Candès EJ, Donoho DL. New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise  $C^2$  Singularities[J]. Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 2004, 57(2): 219-266.
- [24] Do MN, Vetterli M. Contourlets: A New Directional Multiresolution Image Representation[C]//Conference Record of the Thirty-sixth Asilomar Conference on Signals, Systems and Computers, 2002, IEEE, 2002: 497-501.
- [25] Da Cunha AL, Zhou J, Do MN. The Nonsampled Contourlet Transform: Theory, Design, and Applications[J]. IEEE Transactions on Image Processing, 2006, 15(10): 3089-3101.
- [26] Lu Y, Do MN. A New Contourlet Transform with Sharp Frequency Localization[C]//2006 International Conference on Image Processing, IEEE, 2006: 1629-1632.

- [27] Guo K, Labate D. Optimally Sparse Multidimensional Representation Using Shearlets[J]. *Siam Journal on Mathematical Analysis*, 2007, 39(1): 298-318.
- [28] 崔屹. 图像处理与分析: 数学形态学方法及应用[M]. 科学出版社, 2000.
- [29] Schmidt U, Roth S. Shrinkage Fields for Effective Image Restoration[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2774-2781.
- [30] Chen Y, Pock T. Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(6): 1256-1272.
- [31] Yuan Q, Zhang Q, Li J, et al. Hyperspectral Image Denoising Employing a Spatial-spectral Deep Residual Convolutional Neural Network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 57(2): 1205-1218.
- [32] Chen L, Papandreou G, Kokkinos I, et al. Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834-848.
- [33] Cao Z, Hidalgo G, Simon T, et al. Openpose: Realtime Multi-person 2d Pose Estimation Using Part Affinity Fields[J]. *Arxiv Preprint Arxiv:1812.08008*.
- [34] Fang H, Xie S, Tai Y, et al. Rmpe: Regional Multi-person Pose Estimation[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2334-2343.
- [35] Newell A, Huang Z, Deng J. Associative Embedding: End-to-end Learning for Joint Detection and Grouping[C]//Advances in Neural Information Processing Systems, 2017: 2277-2287.
- [36] Hu H, Cai Q, Wang D, et al. Joint Monocular 3d Vehicle Detection and Tracking[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019: 5390-5399.
- [37] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J], *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 34(7): 1409-1422.
- [38] Xu J, Zhao R, Zhu F, et al. Attention-aware Compositional Network for Person Re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2119-2128.

- [39] Wang D, Devin C, Cai Q, et al. Deep Object-centric Policies for Autonomous Driving[C]//2019 International Conference on Robotics and Automation (ICRA), IEEE, 2019: 8853-8859.
- [40] Antol S, Agrawal A, Lu J, et al. VQA: Visual Question Answering[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 2425-2433.
- [41] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [42] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [43] Li Y, Qi H, Dai J, et al. Fully Convolutional Instance-aware Semantic Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2359-2367.
- [44] Lin T, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [45] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-time Object Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [46] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks[C]//Advances in Neural Information Processing Systems, 2015: 91-99.
- [47] Carreira J, Agrawal P, Fragkiadaki K, et al. Human Pose Estimation with Iterative Error Feedback[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4733-4742.
- [48] Wei S, Ramakrishna V, Kanade T, et al. Convolutional Pose Machines[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4724-4732.
- [49] Newell A, Yang K, Deng J. Stacked Hourglass Networks for Human Pose Estimation[C]//European Conference on Computer Vision, Springer, 2016: 483-499.
- [50] Chen Y, Wang Z, Peng Y, et al. Cascaded Pyramid Network for Multi-person Pose Estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7103-7112.

- [51] Peng C, Zhang X, Yu G, et al. Large Kernel Matters--improve Semantic Segmentation by Global Convolutional Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4353-4361.
- [52] Chen L, Zhu Y, Papandreou G, et al. Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [53] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [54] Besl PJ, Mckay ND. Method for Registration of 3-d Shapes[C]//Sensor Fusion Iv: Control Paradigms and Data Structures, International Society for Optics and Photonics, 1992: 586-606.
- [55] Sharp GC, Lee SW, Wehe DK. ICP Registration Using Invariant Features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(1): 90-102.
- [56] Johnson AE, Hebert M. Using Spin Images for Efficient Object Recognition in Cluttered 3d Scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(5): 433-449.
- [57] Mahmoudi S, Daoudi M. 3d Models Retrieval by Using Characteristic Views[C]//Object Recognition Supported by User Interaction for Service Robots, IEEE, 2002: 457-460.
- [58] Wu S, Rupprecht C, Vedaldi A. Unsupervised Learning of Probably Symmetric Deformable 3d Objects from Images in the Wild[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1-10.
- [59] Chen Z, Tagliasacchi A, Zhang H. BSP-net: Generating Compact Meshes Via Binary Space Partitioning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 45-54.



## 版权声明

AMiner 研究报告版权为 AMiner 团队独家所有，拥有唯一著作权。AMiner 咨询产品是 AMiner 团队的研究与统计成果，其性质是供用户内部参考的资料。

AMiner 研究报告提供给订阅用户使用，仅限于用户内部使用。未获得 AMiner 团队授权，任何人和单位不得以任何方式在任何媒体上(包括互联网)公开发布、复制，且不得以任何方式将研究报告的内容提供给其他单位或个人使用。如引用、刊发，需注明出处为“AMiner.org”，且不得对本报告进行有悖原意的删节与修改。

AMiner 研究报告是基于 AMiner 团队及其研究员认可的研究资料，所有资料源自 AMiner 后台程序对大数据的自动分析得到，本研究报告仅作为参考，AMiner 团队不保证所分析得到的准确性和完整性，也不承担任何投资者因使用本产品与服务而产生的任何责任。

顾问：黄 高、李涓子

编辑：叶静芸、李雨晴

数据：赵慧军、袁永强



关注“学术头条”并回复“计算机视觉”下载报告  
报告编号 AITR-20-08