·资助成果·

视频大数据高效表达、深度分析与综合利用

朱文武[1] 段凌宇[2] 田永鸿[2*] 赖剑煌[3] 余 志[3]

1. 清华大学，北京 100084
2. 北京大学，北京 100871
3. 中山大学，广州 510275

[摘 要] 图像视频已占互联网流量和全球数据总量的80%以上，大规模图像视频分析能力的提升越来越成为计算与智能技术融合发展的重要标志，并已成为城市综合治理的重要支撑技术。为突破图像视频大数据处理理论与方法、攻克核心关键技术并实现应用落地，国家自然科学基金委员会—广东省人民政府大数据科学研究中心在2017年启动了“视频大数据高效表达、深度分析与综合利用”重大项目。通过项目的实施，构建了数字视网膜架构视频大数据表征、分析与编码的一体化理论框架，牵头打造我国数字视网膜架构视频大数据处理标准体系，实现了在超算和“云脑”智算平台实现视频大数据分析技术集成验证及在多个城市的实际应用。该项目编号为U1611461，负责人为清华大学朱文武，承担单位包括清华大学、北京大学、中山大学、华南理工大学、浙江大学、上海交通大学、复旦大学、武汉大学、南京理工大学。本项目于2019年12月结题，本文对该联合基金项目的执行情况以及项目成果情况进行了总结。

[关键词] 视频大数据；智慧城市；多元空间

1 项目背景和意义

大数据是关系到未来科技及经济社会发展的重大战略领域，蕴含着巨大的社会、经济和科研价值，因而在近年来引起了科技界、产业界和政府部门的高度关注。在各类大数据中，监控视频是“体量最大的大数据”。相对于文本、语音等数据，图像视频的数据量更大、维度更高，其表达、处理、分析、传输和利用的技术挑战性更大，面临着容量瓶颈、智能瓶颈、融合瓶颈三大难题，具体表现为：(1) 如何突破大量人工标注的瓶颈，实现大规模视觉对象的精准识别与高效关联；(2) 如何实现视频大数据的高效表达与压缩；(3) 如何获得时空关联的视觉对象(人、车、物)的特征鲁棒表达以及协同多摄像机理解大范围场景下行人和群体的行为；(4) 如何协同利用和分析多摄像头网络、社交网络、手机位置等多源信息，实现多尺度态势理解和“见微知著”的群体态势

田永鸿 北京大学博雅特聘教授，博士生导师，2018年国家杰出青年科学基金获得者，兼任鹏城实验室人工智能研究中心副主任、鹏城云脑技术总师。主要研究方向为视觉反演计算、神经形态视觉与视频大数据分析处理，累计发表学术论文200余篇，拥有美/中国发明专利80余项，获国家奖2次、省部级一等奖2次，国际期刊和会议最佳论文奖2次，是首届高校计算机专业优秀教师奖励计划获奖者。兼任国际期刊 *IEEE TCSVT* 编委，IEEE数据压缩标准委员会副主席等。

朱文武 清华大学计算机系教授，信息科学与技术国家研究中心副主任，清华大学人工智能研究院大数据智能研究中心主任，国家973项目首席科学家。IEEE Fellow、AAAS Fellow、SPIE Fellow。CCF多媒体专委会主任。主要从事多媒体网络、多媒体大数据智能等研究工作。曾担任IEEE Transactions on Multimedia (TMM) 主编，现担任TMM指导委员会主席，IEEE TCSVT常务副主编。曾9次获ACM及IEEE等国际最佳论文奖。获2018年度国家自然科学二等奖(排名第1)和2012年度国家自然科学二等奖(排名第2)。

收稿日期：2021-06-17；修回日期：2021-09-10

* 通信作者，Email：yhtian@pku.edu.cn

本文受到国家自然科学基金项目(U1611461)的资助。

预测。因此，存储、传输、处理、分析、应用大规模监控视频数据已成为信息领域面临的重大挑战。本项目聚焦“视频大数据高效表达、深度分析与综合利用”，具有极其重要的理论意义与十分迫切的应用需求。

2　项目执行情况

总体来说，项目组四年来面向国家重大应用需求，立足世界科技前沿的难点，围绕视频大数据高效表达与压缩、视觉对象和事件跨域关联与识别、群体视觉感知与多源异构信息映射三个科学问题开展了深入研究，构建了数字视网膜架构视频大数据表征、分析与编码的一体化理论框架，提出了鲁棒可解释的三元空间视频大数据深层关联表征学习理论方法，突破了编解码协同的千倍率监控视频编转码、万倍率特征表达压缩、多粒度视觉特征表达与高效索引、多源异构大数据表征及跨模态知识图谱构建等关键技术，实现了视频大数据高效压缩和保真表达的有机结合，建立了数据驱动和知识引导相结合的视频大数据计算新模式。在平台建设与应用方面，面向公安与交通深度融合场景构建了基于超算平台的视频大数据分层处理与集成应用环境，在天河二号上实现了百亿幅以上图像大数据深度分析技术验证，在鹏城云脑上实现了万路以上视频大数据处理分析技术验证，在广州、深圳、东莞、韶关等多个城市进行大规模应用示范。

从具体分工来看，在视频大数据高效表达与压缩方面，课题一和课题二分别完成了图像大数据中车辆的实时搜索与精准识别以及十万路视频实时编码分析技术验证；在视觉对象和事件跨域关联与识别方面，课题三实现了多任务视觉特征提取与融合；在群体视觉感知与多源异构信息映射方面，课题四实现了群智感知与多源异构跨媒体融合的视觉大数据分析与预测；在超算平台的技术验证环境与评测体系方面，课题五与课题一和课题二联合实现了超算环境下的超大规模视频图像大数据处理架构；在天河二号超算进行大规模技术验证方面，搭建了千路高清视频编码处理引擎，实现了百亿量级车辆图像高效索引与识别以及十万移动对象的轨迹发现、识别和跟踪。

本项目的实施为城市治理提供了态势感知与预测分析、苗头预判及有序引导的理论基础和技术保障，有效服务了城市综合治理的重大战略需求，取得了显著的应用成效。在项目实施过程中，获得 2017 年国家技术发明奖二等奖、2018 年国家自然科学奖二等奖、2019 年国家科技进步奖二等奖以及 4 个省部级一等奖，牵头制定我国首个人工智能技术国家标准，在有重要影响的国际学术期刊和会议上共发表论文 300 余篇，包括最佳论文奖 7 项，申请发明专利 100 余项。

3　项目研究成果

本项目在理论方法、关键技术、系统应用三方面取得了若干研究成果，具体如下。

3.1　理论方法方面成果

构建了数字视网膜架构视频大数据表征、分析与编码的一体化理论框架，提出了鲁棒可解释的三元空间视频大数据深层关联表征学习理论方法，建立了数据驱动和知识引导相结合的视频大数据计算新模式。

（1）系统性地提出了面向视频大数据高效处理的数字视网膜架构，在率失真理论下建立了视频大数据表征、分析与编码的一体化理论框架[1]。

针对数据量巨大但价值密度极低所导致的视频大数据汇聚分析难题，借鉴“人类视网膜同时具有影像编码与特征编码功能”这一特性，提出了具有“特征实时汇聚＋视频按需调取＋模型增量更新”特性的数字视网膜架构。在率失真理论下依次建模“信号、特征、语义内容”与码率的联合优化模型，提出了面向特征保持的特征率—失准模型和面向目标保留的重要内容次模背包优化模型，建立了视频大数据表征、分析与编码的一体化理论框架。

（2）提出了鲁棒可解释的三元空间视频大数据深层关联表征学习理论方法，构建了多源异构大数据表征及跨模态知识图谱构建方法，建立了数据驱动和知识引导相结合的视频大数据计算新模式[2,3]。

针对视频大数据中物理世界、信息空间和人类社会三元空间跨媒体数据的融合瓶颈，提出了多源异构数据和群智知识计算交互映射的深层关联表征学习理论方法，首次将概率分布表征思想及解耦表征思想引入网络表征学习，设计了基于图卷积神经网络模型的鲁棒、可解释表征学习方法，构建了首个拓扑结构保持的网络深层表征学习模型，解决了有向图表征学习中的非对称传递性难题。通过本项目的研究，建立了拓扑结构与性质保持的关联网络表征学习方法体系。

3.2　关键技术方面成果

突破了一批视频大数据高效表达与深度分析关

键技术，牵头打造我国数字视网膜架构视频大数据处理标准体系。

（1）突破了千倍率监控视频编转码和万倍率特征表达压缩关键技术，建立了基于千万规模车辆的跨场景车辆精准识别框架[4, 5]。

突破了编解码协同的千倍率监控视频编转码关键技术，监控压缩效率比 H.265 标准提升了一倍，建立了全局和局部深度学习特征的联合编码压缩框架，在保持特征性能的同时实现了万倍率特征压缩，提出了面向语义焦点的细粒度视频描述和摘要方法，实现了视频大数据高效压缩和保真表达的有机结合，从根本上解决了视频大数据汇聚分析的技术难题。

针对视频中物体尺度变化与跨模态匹配问题，提出了多层次抽象表示的上下文感知优化模型以及端到端“点—集”匹配深度网络学习模型，大幅提升了图像视频精细识别与分析算法性能，建立了“多粒度视觉特征表达—视觉特征模式发现—大规模视觉特征高效索引”的跨场景车辆精准识别框架，在 22 个基准评测数据集上均获得了性能提升。

（2）牵头制定我国首个人工智能技术国家标准，主导制定视频特征描述子国际标准 CDVA，打造我国数字视网膜架构视频大数据处理标准体系[6]。

项目组全面研发构建我国视频大数据处理核心支撑技术群、测试基准工具链和算法仓，牵头打造我国数字视网膜架构视频大数据处理标准体系：针对视频大数据领域中不同深度学习框架中模型表示不相同、网络复杂度高等难题，提出了神经网络模型的统一表示框架，牵头制定我国首个人工智能技术国家标准《神经网络表示与模型压缩标准》（国标计划号：20192138-T-469）（见图 1），不仅能打破不同 AI 算法框架的壁垒，同时支持神经网络模型的高效压缩来提升资源受限设备和云平台的执行和存储效率；针对视频监控、自动驾驶、智慧城市等应用场景下的视频分析需求，主导制定视频特征描述子国际标准 CDVA。

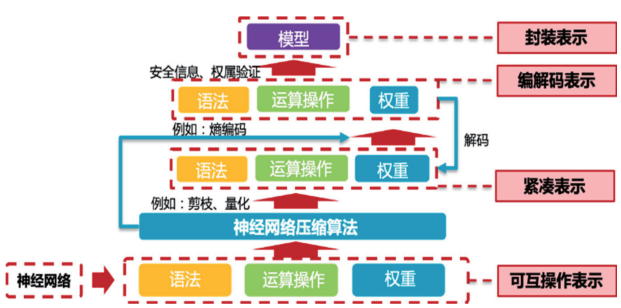

图 1 我国首个人工智能技术国家标准《神经网络表示与模型压缩标准》框架

基于本项目的数字视网膜关键技术研究成果，项目组推动制定科技部新一代人工智能产业技术创新战略联盟的团体标准《信息技术 数字视网膜系统 第 1 部分：系统结构和通信协议》（标准立项号 2020082001）。该标准规定了数字视网膜系统的架构、组成、功能、接口、信息传输流程、通信协议以及安全性要求等内容，适用于数字视网膜系统的方案设计、系统检测、验收以及与之相关的设备研发、生产，其他信息系统可参考采用。

3.3 系统应用方面成果

充分发挥天河二号超算平台与“鹏城云脑”的作用，实现了在超算和云脑超级算力平台实现视频大数据分析技术集成验证及在多个城市的实际应用。

（1）构建了基于超算平台的视频大数据技术集成验证系统，汇聚千万车辆图片、百亿车辆索引、10 亿 GPS 数据配准关联的省域超大规模视频关联数据资源，建立了城市级交通视频大数据综合应用验证平台，实现了多个城市的实际应用。

通过集成各课题核心技术成果，实现了基于天河二号的视频大数据集成开发环境，建成了视频大数据技术集成验证平台（见图 2）。首先，针对视频图像数据特点，建立了“集中—节点”混合分布式存储模式，优化了分布式工作节点间的通信，构建了多模高效计算架构，实现了六网数据联通的研究支撑环境；其次，基于 GIS-T 的视频图像关联存储模型，加载了省域超大规模视频关联数据集，包括详细标注与配准关联的 1 200 万车辆图片、百亿车辆索引、10 亿 GPS 数据等，有效支撑城市级大数据综合业务应用重构和技术评测；第三，通过标准接口接入了 6 项关键技术实现验证，成为典型的城市级视频大数据技术方法和业务应用的综合测试验证平台，进一步重构了“车辆大数据”“视频云＋”等公安实际业务系统，在超算平台上实现了公安实战业务综合应用，在广州、深圳、东莞、韶关等多个城市进行大规模应用示范。

（2）研制了数字视网膜架构的摄像机与阵列服务器，在深圳市光明区完成了千路级高清视频编转码处理，在“鹏城云脑”智能超级算力平台上实现了万路级视频大数据分析处理示范验证，取得显著应用成效。

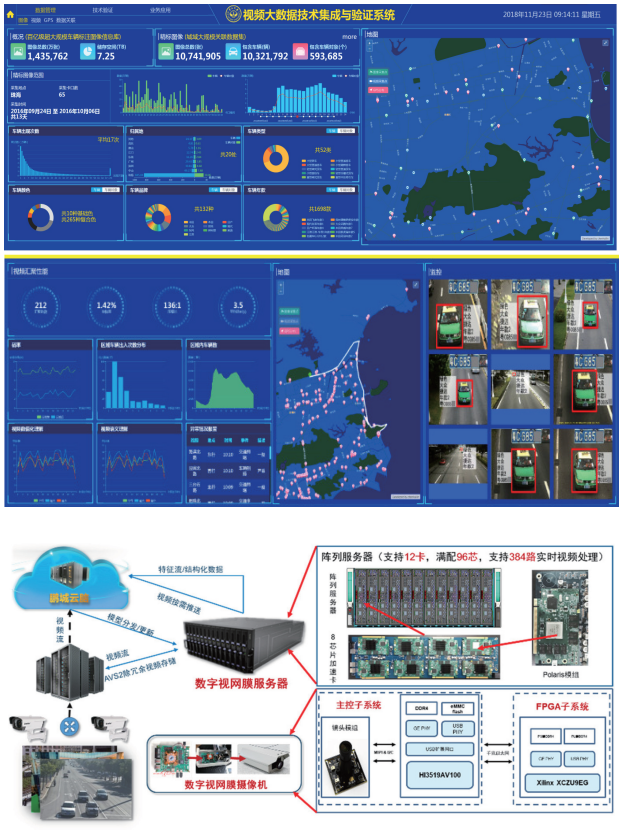

图 2　视频大数据技术集成与验证系统界面

基于国产自主知识产权芯片，研制了数字视网膜架构的摄像机、加速卡与阵列服务器，在“鹏城云脑”智算平台上研发了视频大数据处理分析系统（见图 3），实现了全局统一的时空 ID、高效视频编码和特征编码、功能可定义、模型可更新等数字视网膜核心功能。数字视网膜服务器最大支持 384 路视频分析处理能力，实现了从视频调取、特征汇聚到模型训练分发的闭环。目前，云脑视网膜系统已接入深圳交警 5 500 路视频，并可实时访问监控视频超 1 万路，能高效实现内容压缩、特征提取、车辆与行人目标搜索跟踪，其中千路级高清视频编转码处理的 24 小时视频压缩率达 70％以上，万路级视频大数据中车辆与行人搜索跟踪的平均准确率≥90％。与云天励飞、深圳巴士集团等企业进行合作进行落地应用，利用云脑视网膜系统在深圳市 5 条路 52 个路口上进行了区域级优化测试，可将区域拥堵指数下降 9％，路网协调率提升至 70％，平均车流速度提升 15％，高峰时间缩短 15％，路段平峰平均速度提升 32％，平均行程时间缩减 25.3％，受到深圳交警的高度评价。

图 3　数字视网膜架构摄像机、阵列服务器和视频大数据处理分析系统

4　学术交流、人才培养以及平台建设

4.1　学术交流

在项目执行期间，项目组主要成员赴国内外参加相关研究领域的国际国内学术会议共计 100 余人次，包含 ICML、NeurIPS、CVPR、AAAI 和 ICCV 等国际顶级会议，承办 CIKM2019、MIPR2020、PRCV2018 等学术会议，举办了 CCF 大数据竞赛 2020、2019 和 2020 年全国人工智能大赛。

4.2　人才培养

项目骨干成员 1 人当选欧洲科学院院士、2 人获国家杰出青年科学基金资助、1 人获国家自然科学基金委员会优秀青年基金资助、1 人入选“AI2000”人工智能全球最有影响力学者。

4.3　平台建设

建设了基于超算的视频大数据技术集成与验证系统平台，以及云脑视网膜平台。

5　项目对相关领域的推动作用

在理论方法层面，本项目针对视频大数据“容量瓶颈”“智能瓶颈”“融合瓶颈”三大瓶颈难题，建立了视频大数据高效表达与深度分析的理论体系，提出了分析与编码一体化的视频编码表征理论以及多源异构数据关联表征和群智知识交互映射理论；关键技术方面，本项目提出了视觉对象精准表达的高效索引方法，实现了千万规模车辆的建模和百亿图像高效检索，提出了高效压缩和保真表达的视频编转码方法，实现了千倍率视频压缩和万倍率特征压缩，提出了融合多元时空信息的关联表征学习方法，实现了万级视觉目标的跨域关联与跟踪。应用示范方面，通过本项目的实施，已实现各项关键技术的应用

验证与集成示范系统，在超算平台上实现百亿幅以上图像大数据和十万路以上视频大数据的深度解析和关联分析，面向公安、交通等领域，深度融合超算与业务应用环境，构建新型的视频大数据分层处理与集成应用体系，并在广东省公安系统进行大规模应用验证与示范。为全面提供城市与社会管理态势分析、苗头预判及有序引导的理论基础和技术支持，并服务于城市与社会管理创新的重大战略需求，在国家社会管理主管部门实现落地验证。

本项目研究成果有助于推动支持百亿幅以上图像和十万路以上监控视频深度解析与关联分析的大数据分析处理平台的实际应用，推动视频等非结构化数据分析与处理技术取得根本性的突破。此外，本项目的实施有助于推动我国实现“全域覆盖、全网共享、全时可用、全程可控”的公共安全视频监控建设联网应用。

参 考 文 献

[1] Ding L, Tian YH, Fan HF, et al. Rate-performance-loss optimization for inter-frame deep feature coding from videos. IEEE Transactions on Image Processing, 2017, 26(12): 5743—5757.

[2] Zhu WW, Wang X, Gao W. Multimedia intelligence: when multimedia meets artificial intelligence. IEEE Transactions on Multimedia, 2020, 22(7): 1823—1835.

[3] Zhu WW, Wang X, Li HZ. Multimodal deep analysis for multimedia. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(10).

[4] Bai Y, Lou YH, Gao F, et al. Group-sensitive triplet embedding for vehicle reidentification. IEEE Transactions on Multimedia, 2018, 20(9): 2385—2399.

[5] Feng ZX, Lai JH, Xie XH. Learning view-specific deep networks for person Re-identification. IEEE Transactions on Image Processing, 2018, 27(7): 3472—3483.

[6] 高文，田永鸿，王坚. 数字视网膜：智慧城市系统演进的关键环节. 中国科学(信息科学)，2018，48(8)：1076—1082.

High-efficiency Representation, in-depth Analysis and Comprehensive Utilization of Video Big Data

Zhu Wenwu[1] Duan Lingyu[2] Tian Yonghong[2*] Lai Jianhuang[3] Yu Zhi[3]

1. *Tsinghua University*, *Beijing* 100084
2. *Peking University*, *Beijing* 100084
3. *Sun Yat-sen University*, *Guangzhou* 510275

Abstract Image and video have occupied more than 80% of Internet traffic and global data. The growing capability of large-scale image and video analysis has increasingly become an important symbol of the development of computing and intelligent technology, and has become an important supporting technology for comprehensive urban governance. In order to realize the theoretical breakthroughs, key technological innovations and application implementations for image and video big data processing, the joint Big Data Science Research Center of National Natural Science Foundation of China and Guangdong Provincial Government launched a major project titled “Efficient Representation, In-depth Analysis and Synthesis of Video Big Data” in 2017. The project number is U1611461, and the person in charge is Zhu Wenwu in Tsinghua University. Several research institutes including Tsinghua University, Peking University, Sun Yat-sen University, South China University of Technology, Zhejiang University, Shanghai Jiaotong University, Fudan University, Wuhan University, and Nanjing University of Science and Technology took part in this project, which was completed in December 2019. This article summarizes the implementation process as well as the research outcome of the major project.

Keywords video big data; smart city; multi-space

（责任编辑 刘 敏）

* Corresponding Author, Email: yhtian@pku.edu.cn