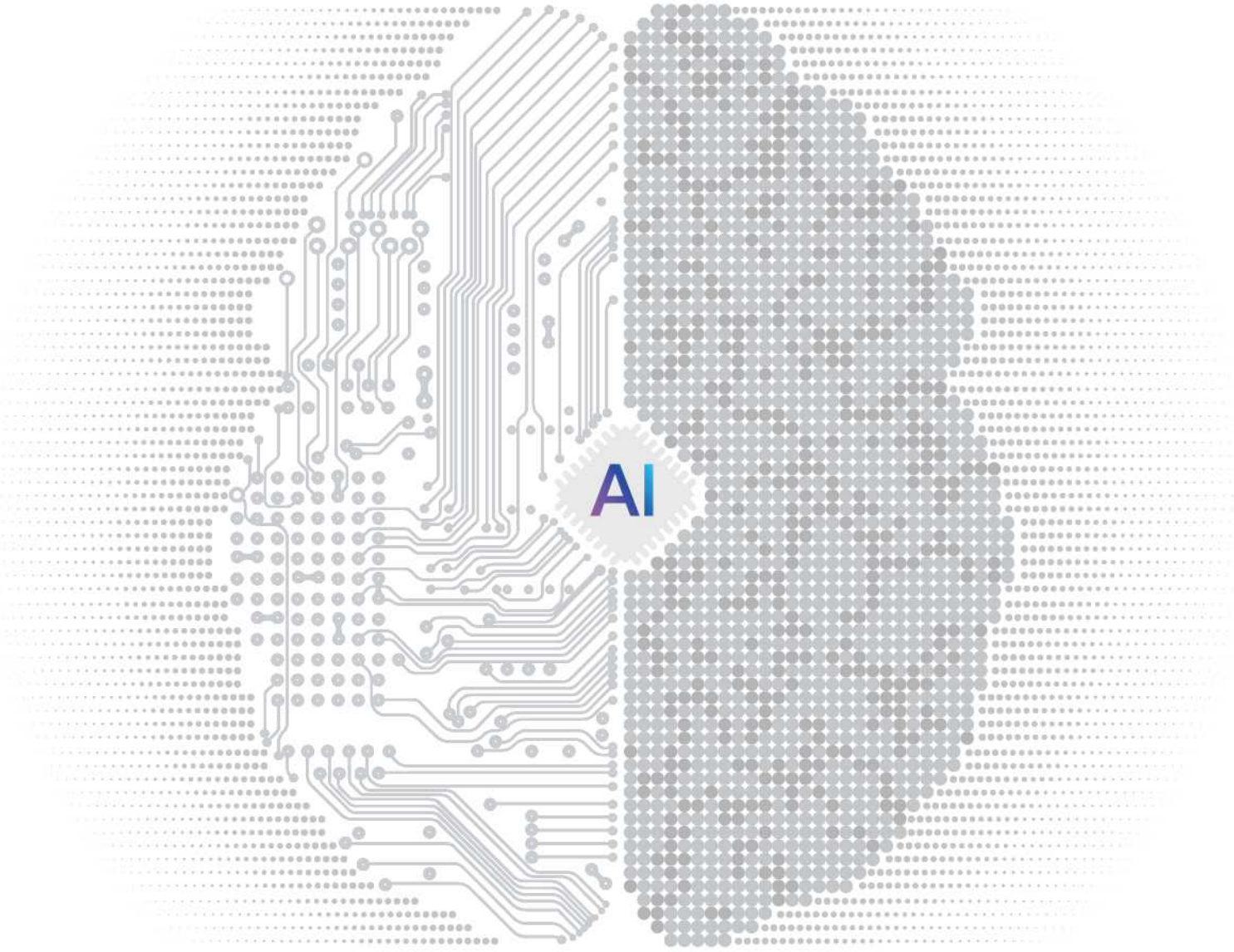
百度大脑AI技术成果白皮书

BAIDU BRAIN AI TECHNOLOGY ACHIEVEMENT WHITE PAPER

2018–2019



Baidu 百度 | 百度研究院

目录

引言

回顾过去的一年，科技与商业发展的一个关键词就是“人工智能”。在近一年的时间里，百度科学家和工程师们不仅在人工智能算法、核心框架、芯片、计算平台、量子计算、语音技术、计算机视觉、增强现实与虚拟现实、语言与知识、开放平台、开放数据等诸多方面取得了令人瞩目的技术成果，还将这些技术成果与行业相结合，成功应用于众多产品之中，取得了丰硕的人工智能应用成果。

2019 年 2 月，世界知识产权组织（World Intellectual Property Organization，简称 WIPO）发布了首份技术趋势报告，聚焦人工智能领域专利申请及发展状况。报告显示，百度在深度学习领域的专利申请量位居全球第二，超越 Alphabet、微软、IBM 等企业和国外学术机构，在全球企业中居于首位。

过去的一年，百度基础技术体系、智能云事业群组和 AI 技术平台体系进行了重大组织机构调整，三个体系统一向集团 CTO 汇报，这为技术中台建设和人工智能技术落地提供了良好的组织保障。

本报告总结了百度大脑在 2018-2019 年度取得的部分技术成果：第一章主要概述百度大脑 5.0，第二至六章分别介绍百度大脑在基础层、感知层、认知层、平台层和安全方面的技术成果。

面向未来，百度将继续打造领先的 AI 技术能力，构建更加繁荣的人工智能生态系统，助力各行各业进入智能化的工业大生产阶段，在智能时代创造更广泛的社会经济价值。

一、百度大脑进化到 5.0

百度大脑是百度AI集大成者。百度大脑自2010年起开始积累基础能力，后逐步完善。2016 年，百度大脑 1.0 完成了部分基础能力和核心技术对外开放；2017 年，2.0 版形成了较为完整的技术体系，开放 60 多项 AI 能力；2018 年，3.0 版在“多模态深度语义理解”上取得重大突破，同时开放 110 多项核心 AI 技术能力；2019 年，百度大脑升级为 5.0，核心技术再获重大突破，实现了 AI 算法、计算架构与应用场景的创新融合，成为软硬件一体的 AI 大生产平台。

如图 1 所示，百度大脑如今已形成了包括基础层、感知层、认知层、平台层以及 AI 安全五大核心架构在内的技术布局。同时，安全一直都贯穿 AI 技术研发的始终，已经融合在百度大脑的所有模块中。基于数据、算法和算力强大的基础能力支持，百度大脑拥有包括语音、视觉、增强现实（AR）/虚拟现实（VR）以及语言与知识等技术能力，并通过 AI 平台对外开放，形成以百度大脑为核心的技术和产业生态。

多年来，百度大脑支持百度几乎所有业务，并面向行业和社会全方位开放，助力合作伙伴和开发者，加速 AI 技术落地应用，赋能各行各业转型升级，其核心技术及开放平台荣获 2018 年度中国电子学会科技进步一等奖。

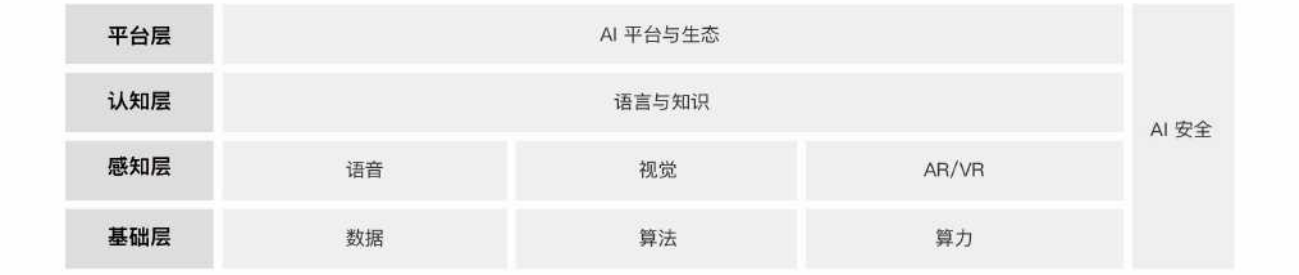


图 1 百度大脑

二、基础层

2.1 算法

百度持续在算法和理论方面深入研究，在语音、图像、语言与知识等多个领域取得重大突破。

在语音识别方面，百度将注意力机制的建模技术用于在线语音识别，提出了流式多层截断注意力模型 SMLTA，实现了流式的基于注意力机制的声学语言一体化建模，并在 2019 年初实现了基于该技术的大规模产品上线，大幅提升了语音识别产品在线识别准确率和用户体验，相对准确率提升 15%至 20%。该算法使用 CTC（Connectionist Temporal Classification）的尖峰信息对连续语音流进行截断，然后在每一个截断的语音小段上进行当前建模单元的注意力建模。通过该方法把原来的全局整句 Attention 建模，变成了局部语音小段的 Attention 建模。同时，为了克服 CTC 模型中不可避免的插入删除错误对系统造成的影响，该算法引入一种特殊的多级 Attention 机制，实现特征层层递进的更精准的特征选择。最终，这种创新建模方法的识别率不但超越了传统的全局 Attention 建模，同时还能够保持计算量、解码速度等在线资源耗费和传统 CTC 模型持平。

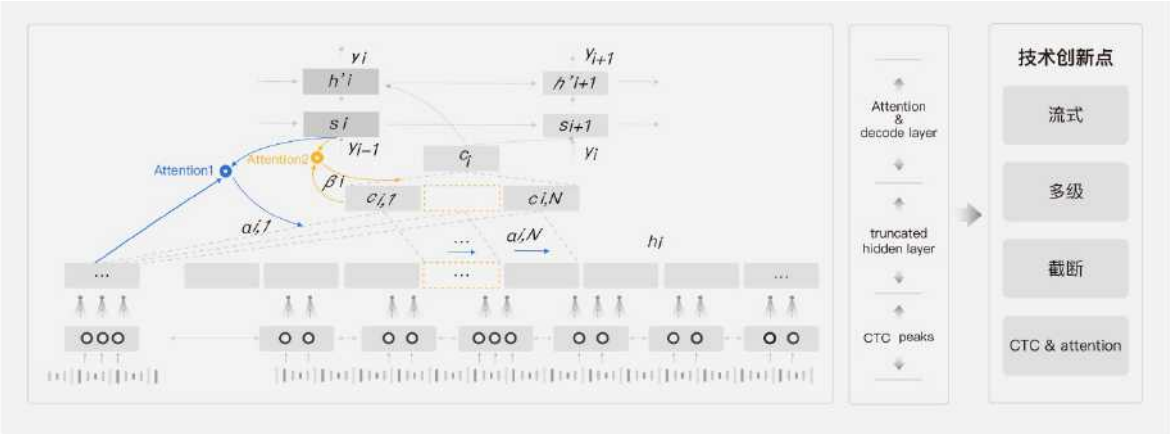


图 2 流式多层截断注意力模型 SMLTA

在个性化语音合成方面，百度还提出了语音风格和音色迁移的个性化韵律迁移语音合成技术 Meitron。该技术在训练时，交叉组合不同声音的训练样本，实现了声音的音色、风格和情感的解耦。语音的个性化信息、风格信息和情感信息等沉淀到全局声音的基

（basis）空间中，并将声音共有信息沉淀到一个统一的声学模型中。在做语音合成的时候，用户仅仅输入少量目标语音作为指导，在全局声音基空间中进行注意力选择，选择出和当前用户个性化声音更加匹配的基。之后可以根据这个基，并结合训练好的共有信息声学模型，生成与目标语音的音色和风格高度相似的任意语音。依靠 Meitron 的解耦和组合机制，我们能够在不同音色、风格和情感之间进行风格转化和迁移，仅仅使用少量用户语音，就可以实现多种音色、情感和风格的转换。该技术成果已经落地百度地图产品，地图用户只需要提供约 20 句话的目标语音，就可以合成与目标语音非常相似的个性化声音，用于地图任意导航场景的语音播报和任意名胜景点的语音播报等。

图 3 Meitron 个性化韵律迁移合成技术

在计算机视觉领域，百度研发了基于图文关系的大规模图像分类弱监督算法，提出了 Ubiquitous Reweighting Network（URNet），给予每张图片训练过程中不同的权重，与原始的分类模型相比，Top5 提升了 8 个点左右。该方法在最大的图像分类数据比赛 Webvision 比赛中获得冠军。在图像超分辨率领域，百度提出了基于级联回归的 CDSR 模型，用于图像的超分增强；还提出了自适应注意力多帧融合技术，用于视频的超分增强。2019 年 5 月，在计算机视觉 Low-level Vision 领域中影响力最大的竞赛 NTIRE 上获得了图像超分辨率项目的冠军和视频超分辨率项目亚军。在医学图像领域，百度提出全新的基于深度学习的病理切片肿瘤检测算法[1]，在公共数据集 Camelyon16 大赛上的肿瘤定位 FROC 分数高达 0.8096，

超过专业病理医生水平以及之前由哈佛、MIT 等保持的大赛最佳成绩。研究成果发表于2018 深度学习医学图像大会。

在自然语言处理领域，百度开发了更具表现力的主题嵌入和知识图嵌入表示学习模型，能够高精度地从语言数据中捕获主题信息。同时，通过联合恢复知识图嵌入空间中的头实体、谓词和尾实体表示，问答系统的回答准确性得到进一步提高。这项工作发表在 IEEE Big Data 2018[2]，SDM 2019[3]，WSDM 2019[4]和 NAACL 2019[5]。

很多高维的特征空间，如词嵌入、图像的特征向量等，都有非常有趣的几何结构。另一方面，多个在语义上有相关性的空间又有一定的相似性。百度深入研究了这些高维空间的特性，提出的全新 Hubless Nearest Neighbor （HNN） Search 算法，能够大幅提高在标准数据集上的单词翻译准确率。以词嵌入空间为例，HNN 能够只用极少量标注数据，实现不同语种间单词的翻译。HNN 此项基础研究能够帮助提升机器翻译系统在低频词、术语、小语种等情况下的效果。另外，HNN 作为一种新的信息检索方法，对广义上的多特征空间匹配都有指导意义，如零样本图像识别等。这项工作发表在 ACL 2019[6]。

百度提出的 Logician 逻辑家代理可以从开放领域自然语言句子中提取事实，实现了更深层次的语言理解，其性能明显优于现有的开放信息提取系统。百度还建立了一个 Orator 演说家代理，可以将几个事实叙述连成一个流利的自然语言句子。通过将提取和叙述作为双重任务，百度在自然语言和知识事实之间搭建了双向的桥梁，使得系统性能得到进一步的提升。这项工作发表在 WSDM 2018[7]和 EMNLP 2018[8]上。

2.2 算力

人工智能时代，算法能力快速提升，同时，算法对算力的要求也越来越高。为了应对算力、效率和多元化场景等核心挑战，百度提出了端到端的 AI 计算架构，通过芯片、连接、系统和调度的协同设计和技术创新，满足 AI 训练方面 IO 密集、计算密集、通信密集的需求，以及 AI 推理方面大吞吐和低延迟的需求。与此同时，包括芯片之间、系统之间、设备之间的互相连接，将帮助不同场景中的计算连接在一起，产生更大的计算力。在系统层兼顾端云，软硬一体，实现了对算力资源的灵活调度。

2.2.1 芯片

云端通用 AI 处理器——百度昆仑

硬件的进展是这次 AI 发展的基础推动力量之一。云端的 AI 推理与训练芯片，成为了各大互联网公司、传统芯片厂商以及创业公司聚焦的战场。业界正在尝试使用特定领域架构（DSA）解决算力及功耗问题。

2018 年开发者大会，百度发布了国内首款云端通用 AI 处理器“百度昆仑”。它基于 XPU 架构，采用 14nm 三星工艺，在 150 瓦功耗限制下，运算性能高达 260Tops，能解决数据中心对芯片的高性能、低成本、高灵活性三大诉求。百度昆仑芯片具备完整的 toolchains，并开放给开发者，与飞桨（PaddlePaddle）实现了深度结合，打造全栈国产技术生态。功能上同时支持视觉、语音、自然语言处理、推荐、无人车等场景，在众多业界深度学习模型上均拥有很好的性能和效率表现；即将量产的芯片在多个模型上实测性能均超过业界主流芯片。

远场语音交互芯片——百度鸿鹄

远场语音交互芯片“百度鸿鹄”变革了传统芯片设计方法，体现了软件定义芯片的全新设计思路。百度鸿鹄拥有契合 AI 算法需求的核内内存结构设计、分级的内存加载策略、依据 AI 算法调教的 cache 设计和灵活的双核通信机制，最终实现了深度学习计算过程和数据加载的高度并行，一颗芯片即同时满足了远场阵列信号实时处理和超低误报高精度唤醒实时监听的需求。

百度鸿鹄可以支持多达六路的麦克阵列语音信号输入；支持百度领先的麦克阵列信号处理技术，即双声道立体声 AEC 消除、声源定位、波束生成等；支持百度领先的 Deep Peak 和 Deep CNN 语音唤醒技术，实现复杂内外噪场景下的高精准唤醒以及低于一天一次的误报率。同时，该芯片还支持百度创新的双麦克模型波束算法，实现唤醒后 360 度无死角识别，首次在中文语音识别上实现双麦克阵列的识别率超越传统 6 麦克系统，实现了行业领先的芯片模型波束技术突破。

2.2.2 AI 计算平台

百度推出的 AI 计算平台，提供了一个端到端的解决方案来应对人工智能计算的挑战。AI 计算平台由超级计算模块 X-Man、高性能存储系统 Fast-F、大型分布式 AI 计算训练平台 KongMing 组成。

X-Man 是百度研发的人工智能超级计算模块，是针对训练场景定制优化的 AI 计算产品。百度在 18 年年底正式发布 X-Man 3.0，单机具备 2000TFlops 算力，并具备灵活的模块化设计功能，能够支持不同的互连架构以及不同的 AI 加速芯片。X-Man 系列产品创造了 6 项业界第一，相关专利荣获了 2018 年中国国家专利优秀奖。百度与 Faceboook、微软等联合创立了 OAI（Open Accelerator Infrastructure）开放 AI 加速基础架构项目，旨在促进 AI 芯片多元化生态格局的健康持续发展。百度在主导 OAI 标准定义的同时，也以实际行动推动 OAI 标准落地，在 19 年 9 月发布了业界首款支持 OAI 标准和液冷散热的超级 AI 计算机 X-Man 4.0。

图 4 百度人工智能超级计算模块 X-Man 4.0

Fast-F 是一种高性能并行文件系统解决方案，硬件上基于 Open Channel SSD 实现 KV 接口，合并 FW 和存储引擎层，软件栈实现全无锁设计，解决了 AI 场景下分布式训练集群中的海量小文件 I/O 难题。

KongMing 是人工智能训练集群，具备自研的高速通信库，充分利用 RDMA 和 NVLink 等特性，并且引入了全网络架构拓扑感知调度，能够以最佳的计算和通信效率将作业映射到多样化的 AI 加速芯片和系统上。KongMing 与 X-Man 及 Fast-F 紧密结合，可支持大规模分布式训练，将训练时间从周级别缩短到天级别。

百度 AI 计算平台已经广泛应用在各行各业的人工智能解决方案中。同时为支撑平台更好地服务业界用户，百度超大规模资源管理系统提供了几十万台服务器托管服务，常驻容器数目达到 500 万，并提供数十万并发计算能力，为大数据处理、模型训练提供支持。

2.2.3 5G 边缘计算

5G 会在许多垂直领域显著提升人工智能服务的能力。近年来，百度一直积极布局边缘计算和 5G 领域。2018 年，百度成功打造出面向互联网的边缘计算统一平台 Over The Edge（OTE），并先后与联通、Intel 等知名企业合作加速 5G 建设。OTE 平台将百度人工智能与 5G 基础设施连接起来，可以使百度人工智能融入万物互联的世界，接近用户，服务用户，成为一个新的生态系统。OTE 平台的架构如图 5 所示，包括资源层的管理，IaaS（Infrastructure as a Service）资源的虚拟化，实现边缘服务管理的 PaaS（Platform as a Service），以及基于 IaaS 和 PaaS 的各种边缘解决方案，可以在边缘提供全面的计算加速支持。

OTE Stack 是面向 5G 和 AI 的边缘计算平台。通过底层的虚拟化，可以屏蔽边缘硬件的异构特性，对外输出标准的算力资源；通过 OTE 层次化的集群管理和全局的智能调度，将 5G 时代大量的边缘节点有效调度起来，从而在边缘为 AI 提供低延迟、高可靠和成本最优的算力支持。同时，通过 OTE Stack 多层集群的统一调度，将设备、移动边缘、云边缘、云中心协同起来，为 Device-Edge-Cloud 的协同计算提供了可能。

图 5 OTE 边缘计算架构

2.2.4 量子计算

量子计算被认为是未来计算技术的心脏。2018 年百度宣布成立量子计算研究所，开展量子计算软件和信息技术应用业务研究，致力于量子信息科学中量子技术的研发和储备，重点关注量子架构、量子算法、以及量子人工智能应用[9][10][11][12][13][14]。

在量子架构方面，百度致力于用半正定规划等优化工具给出任意信道的量子容量可计算上界和信道模拟所需资源估计，这可作为近期量子计算中的量子信道编码、量子纠错和量子电路合成的测试标准。此外，百度探索了量子纠缠这一量子分布式信息处理中最重要物理资源的提纯问题，获得在非渐进（有限资源）情形下的三大参数，即提纯比率、状态拷贝数、以及保真度之间的消长关系。

在量子算法方面，百度利用量子效应设计快速算法来处理非负矩阵分解问题，提供了将量子与经典计算结合起来的“量子分治”策略来加速机器学习的新路径，有望对计算机视觉和机器学习等人工智能应用产生影响。

百度还关注与量子进程有关的问题，回答了“一个量子进程何时比另外一个量子进程更加无序”这一重要问题，从而将著名的优超关系拓展到了量子情形。该关系也给出了量子热力学的一组完整墒条件。

经典算法的改进对于量子计算研究也有极大促进作用。通过改造已有优化算法，百度开发出全新的量子脉冲计算系统“量脉”(Quanlse)，其在量子架构中承接量子软件和量子硬件。对于每一个量子逻辑门，该系统可以快速生成相应的脉冲序列，从而实现对量子硬件的控制。经过实际测试，在相同精度和实验条件下，单量子比特门计算性能比目前最快的工具提升 8 倍以上，而两量子比特门性能则至少提升 23 倍，极大地提升了实验效率。

2.3 数据

过去的一年，百度推出了联邦学习解决方案和数据科学平台等最新成果，并成功运用人工智能技术促进数据工程技术的提升。

联邦学习解决方案

机器学习和深度学习通常需要将数据集中在一个数据中心。近年来，随着整个社会对数据安全及数据隐私的日益重视，以及相关法律法规的出台，使得数据共享和流通面临很多现实挑战。如何在保护数据隐私和数据安全的前提下，利用分散在不同地方的数据来训练机器学习和深度学习模型，成为一个迫切需要解决的问题。联邦学习通过密码学方法和精心设计的模型训练协议，为解决上述问题提供了一种可能的技术手段，能确保隐私数据不出本地的前提下，通过多方协作训练得到一个高精度的机器学习和深度学习模型。

在这个新兴的领域，百度已经设计并实现了针对数据垂直切分场景的分布式 Logistic Regression 联邦学习解决方案，该方案基于参数服务器架构，能够支持在多个节点上并行训练模型，具有良好的可扩展性，可以实现海量数据的联合建模。同时，百度构建了

GBDT 联邦学习的原型系统，并探索了基于深度学习的联邦学习解决方案，包括基于预训练模型的联邦迁移学习以及基于孪生网络结构的联邦学习两类方案。

数据科学平台

百度推出的 Jarvis 数据科学平台，为公司各业务提供易用、高效、自动、安全、节约的统一数据科学环境，大幅提升了开发效率和业务效果，节约大量资源。Jarvis 平台基于 Jupyterlab 的全托管交互分析环境，提供按需弹性的计算资源，成为内部广泛应用的交互环境；基于异构计算的端到端算法加速方案，通过数据科学全流程在 GPU 显存中计算，单机体验好、系统简单易用，分析建模的效率高、成本低，且 GPU 单机比 Spark 集群加速 13 倍，而成本仅为 1/10；支持全流程自动机器学习 AutoML，覆盖预处理、特征工程、模型选择及超参调优等全流程，引入单阶段调优及人工规则优化搜索空间，通过元学习、迁移学习提升搜索效率；支持基于 Jarvis 软件的安全联合建模方案，兼顾安全性和算法效率，保障数据共享、算法分发、建模过程的安全性；支持 GPU 细粒度管理方案，在 GPU 分时复用基础上引入 GPU 卡上计算单位的空分复用，提供任务隔离性和服务质量保证、大幅提升 GPU 资源利用率。

此外，百度在开源社区建设方面也取得突出进展，开源的分布式分析型数据库 Doris 当前在百度以及其他知名互联网公司已大规模使用。在 2018 年进入 Apache 基金会进行孵化后，百度又提供了流式导入功能，对接 Kafka 和增加 SQL 兼容性以及提升查询性能等。

三、感知层

百度大脑的感知层包括语音、视觉、增强现实/虚拟现实等技术，这些技术使得百度大脑具备了仿人的听觉和视觉能力。

3.1 语音

端到端的模型充分发挥了模型联合训练的优势，显著提升了语音识别、语音合成等技术的性能，受到学术界和工业界的一致关注。

语音识别

在流式多层截断注意力模型 SMLTA 的基础上，百度进一步提出了中英文一体化建模和方言大一统模型技术，一方面在保持原有中文识别率的基础上，实现用户中英文混杂和纯英文自由说；另一方面有机融合了普通话和方言的建模单元，使得同一个声学模型既能识别方言又能识别普通话。针对嵌入式终端，百度提出的基于 SMLTA 的离线嵌入式建模技术采用语音语言一体化建模技术，极大地压缩了传统语音识别所使用的语言模型体积。应用该技术的离线语音输入法性能显著领先于行业平均水平。在语音交互方面，百度提出的基于大数据仿真技术的信号前端和声学后端联合优化的整套端到端语音交互方案，使绝对句准率相对提升 3%以上，同时首次推出业内领先的一次唤醒多次交互技术，极大地提升了用户体验。

语音合成

在语音合成方面，百度提出了 End to End Parallel waveRNN（端到端的并行 waveRNN）语音合成技术，解决了语音合成系统上线时的 bad case 消除问题，明显提升了语音合成系统的自然度和表现力，适合大规模在线实时语音合成服务。相比于国际上主流的 Tacotron 和 waveRNN 技术，该技术主要有以下两方面创新：首先，传统方法将文本韵律预测与语音波形生成作为两个独立阶段进行建模学习，由于统计模型误差累积，最终合成语音的 bad case 较多，情感表现力也较弱。端到端的 waveRNN 直接根据输入文本信息，训练一个深度学习 waveRNN 网络以合成语音。整个过程采用端到端训练，不需要中间的梅尔谱的过渡转换过程，减少了合成的 bad case，提升了最终合成语音的自然程度。其次，传统的 waveRNN 是逐点递推过程，递推过程冗长，难以适用于在线实时语音合成的场合。百度能够按照音素、音节、或者音节组合等语音片段单元来独立且并行的合成一个个语音片段，最终再把这些语音片段拼在一起。在进行每个片段的独立合成时候，该片段的 RNN/LSTM/GRU 隐状态的初始状态用传统的拼接系统的决策树提供，从而保证每个独立合成片段的合成稳定性。

百度还提出了一种全新的基于 WaveNet 的并行音频波形（Raw Audio Waveform）生成模型 ClariNet[15]，合成速度比起原始的 WaveNet 提升了数千倍，可以达到实时合成速度的二十倍——即合成 1 秒语音，只需要 0.05 秒。ClariNet 是语音合成领域第一个完全端到端

的模型，即使用单个神经网络，直接从文本输入到原始音频波形输出。对比 Google DeepMind 提出的 Parallel WaveNet，ClariNet 中所用到的 teacher WaveNet 的输出概率分布是一个方差有下界的单高斯分布，直接使用最大似然估计来学习，并且直接闭式（colsed-form）地计算目标函数，大大简化了训练算法，使训练时间比 Parallel WaveNet 减少数十倍。

另外，百度提出了针对语音合成领域的全并行模型 ParaNet[16]。该模型直接采用前馈神经网络（Feedforward Neural Network），不依赖于任何自回归神经网络（Autoregressive Neural Network）或者循环神经网络（RNN），从文本生成音频波形仅需一次前馈传导（Feed-Forward Pass），合成速度较全卷积的自回归模型提升了 46.7 倍。在长句的合成过程中，ParaNet 提供了更为稳定的文本与频谱之间的对齐关系，减少了重复词、跳词、以及错误发音，相比于自回归模型有更高的鲁棒性。

3.2 视觉

在计算机视觉方面，百度在基础图像技术、视频理解技术、软硬件结合等多个技术方向，取得了重要突破，多次获得顶级赛事的冠军。

图像技术

目标检测是计算机视觉和模式识别领域的基础问题之一，百度在大规模图像检测和检测网络的性能两个方面开展研发工作。

一方面，在图像基础算法方面，百度研发了大规模图像检测算法。该算法提出的动态采样方案，对于不同类别，数据量差别较大的情况下效果有明显提升。2018 年，百度在图像检测数据 Open Images 比赛中获得全球第一名的成绩。该技术并已被应用于商品检索、Logo 检索等多个业务中。

另一方面，百度在图像检测领域获得 2019 年“Objects365 物体检测”国际竞赛 Full Track 冠军。在这次比赛中，百度通过使用基于强化学习的网络结构搜索技术，大幅增强了 Two-Stage 检测网络模型的性能，并针对大规模图像检测任务提出的 Class Aware Sensitive 采样方案，有效的缩短了模型收敛所需的迭代次数，进一步提高了模型的最终效果。

视频技术

百度视频理解技术持续优化，支持百度搜索的视频数据分析的相关业务需求。目前小视频分类准确率超过 90%，业务上优质视频增益率达 95%以上。百度连续三年在视频理解领域影响力最大的赛事 ActivityNet 上获得冠军。

在视频编辑方面，百度结合多模态嘴型生成、GAN、TTS 等技术，实现了业界首个可以量产视频的真人形象虚拟主播，并成功应用于好看视频天气预告、新闻播报等场景。百度还提出了选择性迁移单元技术用于提升 GAN 的表现效果，在公开数据集 CelebA 取得了世界领先的效果，相关算法 STGAN 的工作内容发表于国际顶级学术会议 CVPR 2019[17]。

图 6 百度真人形象虚拟主播

在人体视觉理解方面，百度对以往基于多尺度全卷积神经网络的模型（例如 Pyramid Scene Parsing Network，DeepLab v3+等）进行改进，使每个卷积核能对图片的细节进行感知，同时输出精度更高的特征图，解决了人体关键目标区域较小，难以检测的问题。此外，百度还进行了图片增强、数据扩张，在训练中动态调整输入图片尺度，使用 mIOU loss 损失函数等，使得模型能够更精确地捕捉肢体的细节以及被遮盖的部分。最终根据各个不同模型的效果进行融合，在 CVPR 2019 LIP（Look Into Person）竞赛中，百度取得 65.18%的 mIoU，获得了单人人体解析的冠军，超过上届冠军 7.2 个百分点，总计获得人体检测领域三项冠军。

在智能城市视频分析领域，百度参加了由 NVIDIA 在 CVPR 2019 上举办的 AI CITY 比赛，并拿到车辆 RE-ID 第一。在车辆重识别技术方面，百度深耕检测、跟踪、属性分析、关键点定位等核心技术，设计出基于关键点的特征图变换网络，并结合车型识别、摄像头时空分布信息等辅助手段提升车辆重识别准确率。这项技术广泛服务于城市安防、智能城市、智能交通等重要的 AI2B 场景。

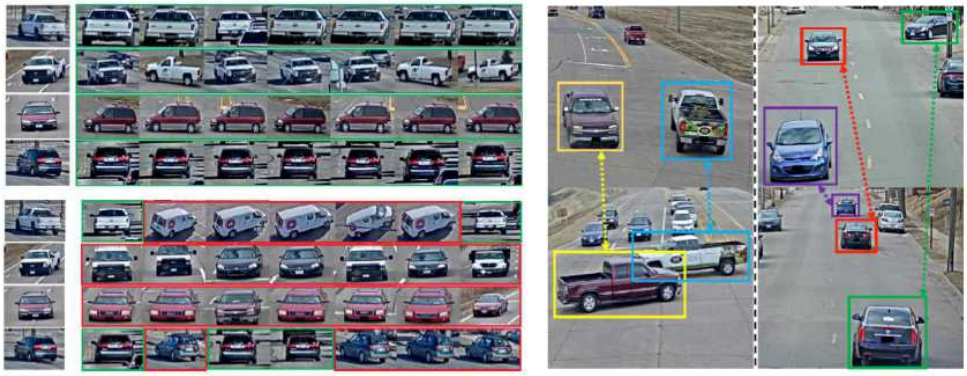


图 7 百度车辆识别技术效果

在视频跟踪方面，百度还在多尺度特征提取、改善物体模板以提升对微小目标的召回能力、利用时空特征来降低密集多目标跟踪的轨迹交换等方面，取得重要进展，并在国际权威的视频多目标追踪挑战（Multiple Object Tracking Challenge，MOT）的 MOT16 榜单上，获得第一名。这些视频能力对内支持百度智能城市、智能零售、自动驾驶等业务，并通过百度大脑 AI 开放平台对外开放，服务各行各业。

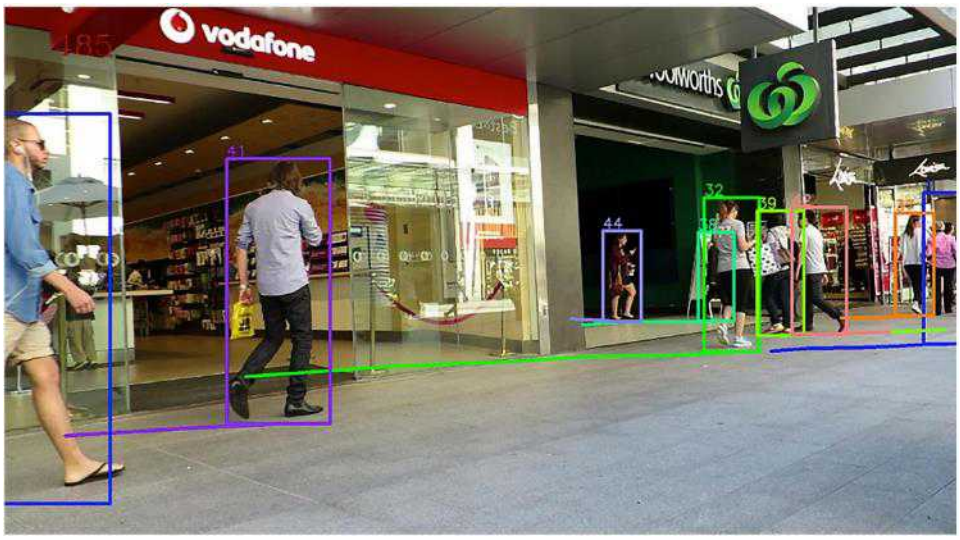


图 8 百度多目标追踪效果

人脸检测与识别

在人脸技术方向，百度在 PyramidBox 基础上提出的 PyramidBox++算法在国际权威评测集 Widerface 最难的 Hard 子集上排名世界第一[18]；在人脸关键点技术上，百度应用 AutoDL 技术取得了 ICME 2019 人脸关键点比赛的冠军。在 CVPR 2019 首次举办的人脸活体检测比赛中，百度作为邀请参赛团队取得了 ACER 指标第一名的优异成绩，即平均错误率最低。这项技术也作为百度 FaceID 解决方案的一项重要功能在多个人脸场景里得到了应用。

文字识别

在文字识别 OCR 领域，百度在端到端文字识别任务上取得了 RCTW-17 世界第一的成绩。在文字检测、结构化文字信息提取、视频 OCR 等多个研究方向上百度也持续探索，相关的成果在 CVPR 2019 和 ICDAR 2019 发表，并在 ICDAR 2019 MLT 多语种文字检测竞赛中，取得了第一的优异成绩。基于在文字识别领域多年的研究和应用经验，百度联合学术界举办了 ICDAR 2019 的两项文字识别竞赛：LSVT（Large-scale Street View Text with Partial Labeling，大规模弱标注街景文字）、ArT（Arbitrary-Shaped Text，任意形状场景文字），吸引了世界范围内高校、知名企业等 100 多支队伍参赛，在弱监督文字识别和任意形状文字识别两个新任务方向上为学术界提供了有力的研究数据和工具。

软硬结合

在视觉模型小型化技术方向，百度形成了一套从模型压缩到模型自动搜索比较完备的体系，囊括了量化、减枝、蒸馏、模型自动压缩、模型自动搜索、硬件搜索等方面，已经在视觉各项任务上得到应用。面向硬件的模型加速，百度研发了 Leopard 系统。该系统综合利用模型压缩、动态训练策略、以及并行化训练架构，实现视觉识别模型的训练推理的显著加速。这项技术在斯坦福大学举办的 DAWN 竞赛（Data Analytics for What's Next）中共取得 CIFAR10 推理速度和成本，以及 CIFAR10 训练速度和成本四项第一。

在实际应用中，视觉语义化往往依赖大量传感器的综合信息，并需要大量计算资源和融合推理的难题，百度研发了多传感器视觉语义化技术。依托边缘视觉计算技术、多种类型自研传感器、以及高 SLA 软件架构，可以实现多人复杂任务的视觉语义化推理。以一个便利店环境为例，百度安装了超过 1000 个多种类型传感器（重力、光幕、相机、深度相

机）。依靠端云结合的人体追踪、肢体检测、商品取放检测、SKU 分类、融合推理等算法，可以在 4 平米/人的密度下准确追踪和分析消费者购物行为，即使多人靠近同时拿取相邻商品也可以准确分辨。为了提升系统运行速度，百度利用 300 多个端计算芯片分担服务器计算负担，减少了 95%的网络传输和 GPU 服务器需求。该技术能够支持更多的单位面积购物人数以及单位面积 SKU 数量，并且可以更快速的完成视觉语义化推理并推送账单。

在机器人避障技术方向，其难点在于检测障碍物的同时需要对自身准确定位，并判断可通过空间的大小。百度提供了市场上领先的机器人视觉 SLAM 定位技术（boteye）；并进一步开发了技术领先的机器人避障技术，利用强化学习算法模型有效融合视觉和激光传感器，以及端到端输出底盘控制信号，提升避障成功率。相比 ROS，在多个场景下，百度机器人避障技术的避障成功率均大幅领先。

3.3 增强现实/虚拟现实

2018 年以来，百度在增强现实和虚拟现实方面取得了许多重要进展。百度构建了生态开放平台 DuMix AR，开放多种 AR 核心能力和 AR 引擎，为开发者及合作伙伴提供优质的一站式解决方案。百度 VR 已在全景、3D 图像内容的采集、处理、传输、展示及交互技术形成了深厚积累。百度还开发了一种增强现实的自动驾驶仿真系统。

增强现实

百度大脑 DuMix AR 平台作为百度大脑的重要组成部分之一，目前已成为国内最具影响力的 AR 技术开放平台之一，累计开放技术能力超过 40 项。最新发布的 DuMix AR 5.0，带来人机交互和感知跟踪两个方向的重大升级。人机交互方面，百度打造人脸人体手势环境一体化交互系统，为业界及合作伙伴提供优质的一站式娱乐互动解决方案。感知跟踪方面，百度自主研发视觉定位与增强服务 VPAS，通过离线高精地图构建、在线定位、融合跟踪等三大关键模块，构建了国内首个达到商用落地标准的大场景物理世界交互系统。

DuMix AR 平台联合 40 多个生态合作伙伴在品牌营销、视频娱乐、景区、教育和汽车等多个垂直行业开展创新探索。继 AR 太极大屏落地全球首个 AI 公园——海淀公园后，2019 年，AR 太极大屏迅速推广至全国多个城市，丰富线下互动体验、引发全民健身热潮，累计落地九个公园，十五块 AR 互动屏，并衍生出八段锦等创新互动内容形态；2019 年，春晚

切红包、虚拟主播“小灵”先后登陆央视；“听障儿童无障碍阅读计划”携手百度公益、壹基金、58 同城，以 AR 技术变革传统出版物，关爱弱势群体，创造良好的社会效益；此外，还与百度地图场景化能力结合，率先实现大场景实景 AR 互动，以 VPAS 再现圆明园大水法的辉煌盛景，将历史画卷生动呈现。截止目前，DuMix AR 平台承载的累计互动量超过 19 亿，深入 6 大行业发布解决方案，显著推动了 AR 技术与应用的发展。

虚拟现实

百度也在深耕 VR 核心技术和行业落地解决方案。在全景、3D 图像内容的采集、处理、传输、展示及交互技术持续积累，已支持 K12、高校培训、营销等业务场景的落地。在内容生产方面，通过高精度拍摄、智能拼接、基于深度学习的图像精准分割技术，构建了软硬一体化的 3D 图像采集方案；在内容展示方面，基于自研 webVR 渲染引擎、自研高性能全景和 3D 图片视频播放引擎以及长期积累的 VR 头显硬件适配能力，推出了可覆盖 Unity、Web 平台的全格式内容展示 VR Suite SDK，为合作伙伴提供基础的 VR 内容播放技术支持。目前，教育方向产品“百度 VR 教室”已经在安徽、浙江、上海、湖北、天津等多地学校落地并常态化运营；“VR 新商科实验室”也已在山大、矿大多所高校落地中。百度还积极推动 VR 技术在 5G+教育场景有效落地，上海市愚一小学已成为全国首家应用 5G Cloud VR 的教学点。营销方面，百度 VR 联合优信二手车，推出了全国领先的“VR 看车”软硬一体化解决方案，助力优信二手车全国购战略升级。

增强现实的自动驾驶仿真

自动驾驶系统对安全性有着严苛的要求，相比于花费几年甚至更久时间让自动驾驶车辆接受足够的道路测试，通过仿真系统测试来评估、提高其安全性极具可操作性和现实意义。百度开发了一种增强现实的自动驾驶仿真系统，通过模拟交通流来增强现实世界图像，进而创建逼真的、媲美现实世界渲染的仿真场景，为自动驾驶车辆提供更为可靠且廉价的仿真模拟方法，可大规模用于训练和测试评估自动驾驶系统的感知、决策和路径规划算法。该系统相较于现有仿真系统，在真实感、扩展性等方面都实现了突破性的技术进展，并发表于《科学》杂志《机器人学》子刊[19]。

四、认知层

多年来，百度深耕语言与知识技术，并在知识图谱、语义理解、机器翻译等方面取得了一系列丰硕的成果，实现了大规模产业化应用，获得国家科技进步二等奖、中国电子学会科技进步一等奖、中国专利银奖等奖励。

百度知识图谱依托海量互联网数据，综合运用语义理解、知识挖掘、知识整合与补全等技术，提炼出高精度知识，并组织成图谱，进而基于知识图谱进行理解、推理和计算。目前，百度知识图谱已经拥有数亿实体、数千亿事实，并广泛应用于百度众多产品线，并通过百度大脑 AI 开放平台开放了实体标注、知识问答、百度汉语、图数据库 BGraph 等核心技术，以及行业知识图谱平台和智能写作平台。百度研发了基于深度学习的语义理解技术并应用于智能搜索，大幅提升了搜索精度；百度提出了持续学习语义理解框架艾尼（ERNIE），在中英文多项任务上均取得最好的效果；百度机器翻译在大规模翻译知识获取、翻译模型、多语种翻译技术等方面取得重大突破，在 2015 年发布了全球首个互联网神经网络翻译系统，在 2018 年推出了端到端同传系统和翻译机；百度在基于多文档校验的阅读理解技术、基于交互式学习的对话理解技术、篇章生成算法等方面取得突破性创新，在 AI for Prosthetics Challenge、国际语义评测比赛（International Workshop on Semantic Evaluation）、国际机器翻译比赛（WMT）等国际权威赛事中屡获冠军。

智能搜索	深度问答	对话系统	智能创作	机器翻译
语言理解				语言生成
词法	句法	语义	篇章	篇章生成
专名识别	依存分析	逻辑推理	对话理解	摘要生成
词性标注	短语分析	语义计算	情感分析	
分　词	组块分析	语义表示	主题识别	片段提取
知识图谱				
实体	关注点	事件	行业知识	地点
知识挖掘		知识整合与补全		分布式图索引及存储计算

图 9 百度语言与知识技术布局

4.1 知识图谱

知识自学习

知识自动学习和更新是开放域大规模知识图谱构建面临的主要挑战。百度研发了一套自下而上的基于主动学习的大规模知识图谱构建方法，包括开放信息抽取、本体自动构建、图谱自动补全、多源数据融合，以及人机结合的知识验证等多项核心技术。百度提出了基于远监督学习训练语料构建的信息抽取技术，基于 Bootstrapping 算法和深度学习的本体自动构建和图谱补全技术，并通过多源数据融合和人机结合进行知识验证，进一步提升知识获取质量。该方法实现了本体层及数据层知识的自我学习完善及更新，大幅提高了知识图谱构建效率。基于该方法，百度知识图谱扩大了几个数量级，显著提高了百度知识图谱的覆盖率。

复杂知识图谱

传统的知识图谱以实体为核心，被称为实体知识图谱，不能很好地描述行为、状态、时序、空间、条件、概率等复杂知识，尚需研发更强语义表达能力的知识图谱。百度研发了包含事件检测、事件抽取、事件关系分析等核心技术的全流程事件图谱构建方案。基于深度神经网络模型，优化了事件检测的事件判别、事件名生成、事件归一等关键策略，并使得事件发现到收录入库达到分钟级别。在事件本体自动挖掘中，自动挖掘了数千个事件本体，覆盖了大部分的事件类型。具备事件时间、地点、触发词、核心参与者等事件属性抽取能力，同时基于事件时序关系抽取技术能高效地自动挖掘事件脉络。目前已经形成了千万级别的事件图谱，支持了事件脉络、行业热点、地域热点等多个实际场景的应用。此外，基于实际业务需求，百度还构建了关注点图谱，POI（Point of interest）图谱等多元知识图谱，研发了多元知识图谱的关联技术。基于多元知识图谱的丰富表达，支持了众多智能化的产品及应用。

行业知识图谱

近年来，越来越多的行业、企业希望利用知识图谱，沉淀行业知识，提升行业应用的智能水平。百度在行业知识图谱方向重点建设了三个层次的能力。首先，建设了行业知识图谱构建平台，将多年积累的通用知识图谱构建全流程的策略、算法、架构能力迁移到行

业，并针对行业图谱构建定制化高、数据质量参差、启动成本高等特点进行了系统性的设计和升级。其次，建设行业知识图谱应用能力组件，建设了包括行业知识问答，基于BGraph 的图谱检索计算引擎等行业应用强需求的基础能力组件。第三，研发了行业知识图谱解决方案级产品，在智能客服、智慧司法、智能金融和智能企业信息管理等领域打造了一体化解决方案并实现落地。此外，在医疗知识图谱构建和应用方面，通过与医学专家合作，构建了专业的医学本体，并自动构建了百万级的医学实体及千万级医学事实。百度基于专业医疗知识图谱实现了一整套医疗认知计算引擎，提供包括可循证辅助诊断、治疗建议、病历质控、合理用药、病历语义推荐与检索、知识查询、多轮分诊等在内的临床辅助决策能力，并已落地赋能医疗机构，提升诊疗效率与质量。

基于知识图谱的视频语义理解

百度提出了一种基于知识图谱的视频语义理解技术，充分利用知识图谱丰富而全面的事实提升视频语义理解效果。如图 10，该技术通过一种基于多粒度跨模态注意力机制的语义网络实现对视觉、语音、文本的多模态内容解析融合；利用知识子图关联技术建立与视频理解知识图谱的连接，通过背景知识以及基于多模态知识的计算与推理，实现对视频的深度语义理解；突破了传统基于感知技术的视频内容分析的局限，实现对视频的深度结构化的语义理解，大幅提升视频理解的效果。目前已在百度视频搜索、推荐、内容生成等多个产品上线。其中，多粒度跨模态注意力机制的视频理解工作在发表在 ACL 2019[20]。

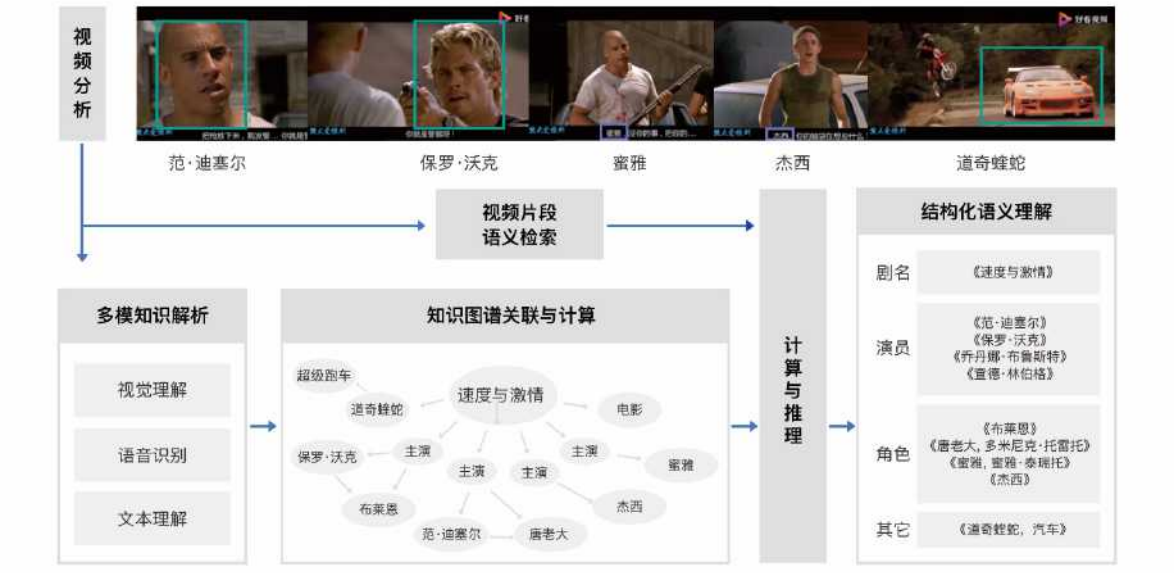


图10 基于知识图谱的视频语义理解

4.2 自然语言处理

语义理解

百度提出的持续学习语义理解框架艾尼（ERNIE）充分利用了百度海量数据和飞桨多机多卡高效训练优势，创新地融合大数据及知识，持续地构建词法、句法和语义三个层次的多种预训练任务，通过持续的多任务学习技术进行训练更新，如图 11 所示。目前，ERNIE 已累计学习超过 13 亿多的知识，全面刷新中英文 NLP 任务效果。

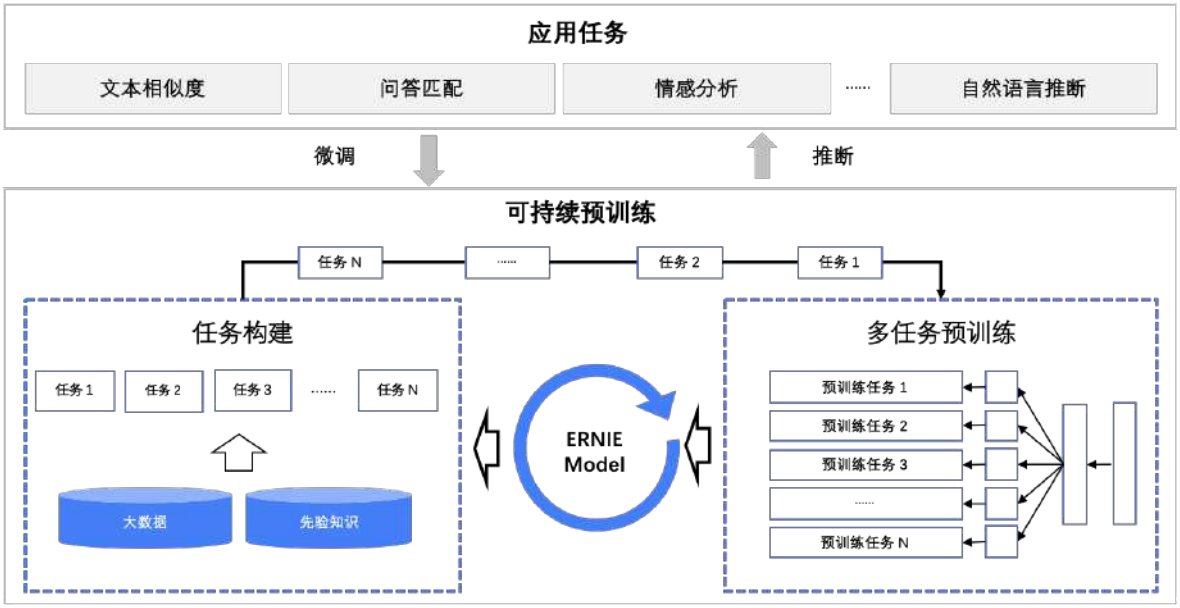


图 11 持续学习语义理解框架 ERNIE 2.0

百度 7 月发布的 ERNIE 2.0 进一步刷新中英文 NLP 任务效果。目前，ERNIE 2.0 在自然语言推断、自动问答、阅读理解等多种中文 NLP 任务和英文通用自然语言理解任务 GLUE 上，效果超越 BERT 和 XLNET，具体如表 1，表 2 所示[1]。百度近期还发布了面向企业应用的 ERNIE 开源工具集，包括功能全面的 Fine-tuning 工具、速度领先的预测工具、灵活的部署工具和一键式的压缩工具等。

[1] 表 1、表 2 表示 2019 年 7 月的测试数据，效果还在持续提升中。

表 1 中文 NLP 任务效果

中文NLP任务	数据集	ERNIE-Large效果	ERNIE-Base效果	BERT-Base效果
自动问答	NLPCC-DBQA	85.8% (+5.0%)	85.3% (+4.5%)	80.8%
自然语言推断	XNLI	81.0% (+3.8%)	79.7% (+2.5%)	77.2%
情感分析	ChnSentiCorp	95.8% (+1.5%)	95.5% (+1.2%)	94.3%
文本语义相似度	LCQMC	87.9% (+0.9%)	87.9% (+0.9%)	87.0%
命名实体识别	MSRA-NER	95.0% (+2.4%)	93.8% (+1.2%)	92.6%
机器阅读理解	Dureader	64.2% (+4.7%)	61.3% (+1.8%)	59.5%
机器阅读理解	CMRC2018	71.5% (+5.2%)	69.1% (+2.8%)	66.3%
机器阅读理解	DRCD	89.0% (+4.1%)	88.0% (+3.1%)	84.9%

表 2 英文 GLUE 数据集合效果

英文NLP任务	数据集	ERNIE-Large效果	BERT-Large效果	XLNET-Large效果
语法判断	CoLA	65.4% (+4.8%, +1.8%)	60.6%	63.6%
情感分析	SST	96.0% (+2.8%, +0.4%)	93.2%	95.6%
语义等价判断	MRPC	89.7% (+1.7%, +0.5%)	88.0%	89.2%
文本语义相似度	STS-B	92.3% (+2.3%, +0.5%)	90.0%	91.8%
语义等价判断	QQP	92.5% (+1.2%, +0.7%)	91.3%	91.8%
自然语言推断	MNLI	89.1% (+2.5%, -0.7%)	86.6%	89.8%
自动问答	QNLI	94.3% (+2.0%, +0.4%)	92.3%	93.9%
蕴涵关系识别	RTE	85.2% (+14.8%, +1.4%)	70.4%	83.8%

机器翻译

百度在2015年发布了全球首个互联网神经网络翻译系统，突破了大规模数据训练、集外词问题、漏译问题、数据稀疏等一系列国际公认难题[21][22]，翻译质量提升幅度超过过去十年总和。

过去的一年，百度结合语音技术、机器翻译技术，研发了全球首个高质量、低时延的端到端同传系统，为用户带来沉浸式同传体验。在语音容错方面，百度提出了联合词向量编码模型[23]，提升模型鲁棒性，缓解语音识别错误；在实时性方面，百度提出了具有预测和可控时延的翻译模型[24]；在语篇翻译方面，百度提出了基于多轮解码校对网络的语篇翻译模型[25]，以解决译文一致性和连贯性问题；在端到端机器同传模型方面，百度首次提出了基于知识蒸馏的同传模型[26]，可以有效克服数据稀疏问题，显著提升翻译质量。

百度与途鸽合作研发了全球首个共享 Wi-Fi 翻译机，创造性地将 Wi-Fi 和语音翻译集成，实现了 80 多个国家的 4G 网络高速连接和 10 种语言的高质量语音翻译。研发了自动语言识别技术，首次实现了一键互译。相比于传统的翻译机，极大地提升了用户体验和交互效率。凭借语音翻译的一系列技术突破，在美国权威杂志《麻省理工科技评论》（MIT Technology Review）2018 年“全球十大突破性技术”（10 Breakthrough Technologies 2018）官方榜单中，百度被列为实时语音翻译领域“关键玩家”，成为本年度唯一一家入选的中国公司[27]。2018 年 8 月，百度翻译机作为官方合作伙伴，支持“中非合作论坛北京峰会”的多语言翻译。

百度翻译支持 28 种语言、756 个翻译方向，覆盖全球 47 亿人口，每日实时响应超过千亿字符的翻译请求，同时，提供了开放云接口服务，支持超过 25 万个第三方应用，极大地促进了相关产业发展。

图 12 百度-途鸽 Wi-Fi 翻译机

可定制对话技术

基于深厚积累的词法分析、语义相似度计算、纠错改写、情绪识别、关键词分析等基础 NLP 技术，百度建立了包括可定制的对话理解技术、问答技术、对话式理解辅助技术等多项核心技术，以及由对话管理、需求分发、全局记忆等机制构成的对话系统框架。

对话理解过程以基于组合语义推导的启发式模型、基于 ERNIE 语义理解框架的深度学习模型，以及多引擎融合技术为支撑，兼顾了对话理解模型的快速启动和深度优化诉求，在典型的办公场景中可以将同等理解效果下的数据需求量降低 30%~70%，总体开发成本降

低 60%。同时，百度还提出了基于深度注意网络的多轮响应选择匹配模型 DAM（Deep Attention Matching Network），显著提高了口语理解能力。此外，百度创新地提出了对话式理解辅助技术，基于异常对话监控与用户反馈识别技术，在与人类的交互过程中优化对话理解、问答的模型效果，可以在五轮对话之内，将理解准确率提升至 98%以上，进一步降低了系统的持续优化成本。

在对话系统框架中，百度一方面提供了可编程的对话管理框架，并内置了多个常用标准对话范式，为在云端开发灵活可变的业务对话逻辑提供了便利，减少了系统集成与接入的开销。另一方面，百度提供了需求分发和全局记忆机制，支持多个对话任务的集成与联动，提高了对话技能的可复用性，降低了新业务的重复开发成本。

百度可定制对话技术依托百度大脑 UNIT 3.0 平台，支持数万多个对话技能，在百度内外取得了广泛应用，包括百度地图语音助手、百度 AI 输入法、百度车联网开放平台等项目，以及中国联通、招商银行、南方航空、东方航空等多个大型行业客户。

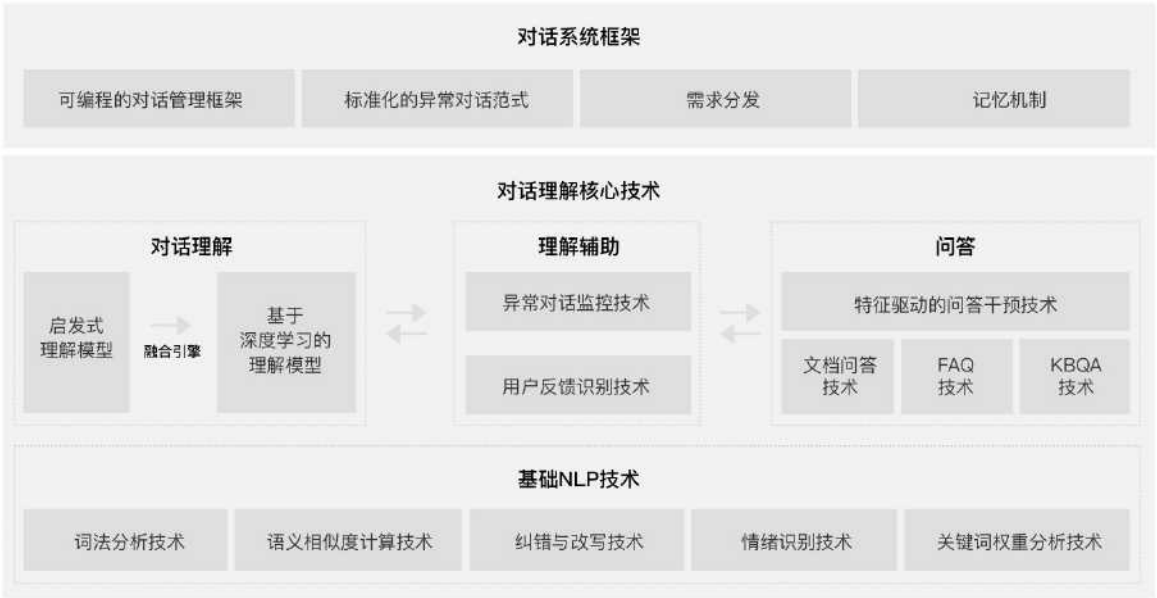


图 13 可定制对话技术系统

智能问答与阅读理解

近年来，基于机器阅读理解（Machine Reading Comprehension）的问答技术受到学术界和工业界的广泛关注。百度提出了一种知识表示和语言表示深度融合的模型 KT-NET，同时借助语言和知识提升机器阅读理解的效果。该模型首先利用注意力机制从知识图谱中自动筛选并整合与阅读内容高度相关的知识，然后通过双层自注意力匹配，实现文本表示和知识表示的深度融合，提升答案边界预测的准确性。截止论文发表，该模型在常识推理阅读理解数据集 ReCoRD 和斯坦福数据集 SQuAD 1.1 均取得了单模型最好成绩[28]。此外，百度探索了一种借助图神经网络（GNN）更好地利用知识图谱表示改进机器阅读理解的方法。通过图注意网络（GAT）对知识图谱中对应的子图结构进行信息整合，为机器阅读理解过程中的决策判断提供更丰富的指导信息，取得了显著的效果提升。

在面向搜索场景的多文档阅读理解中，针对多个候选文档中歧义信息较多的挑战，百度研发了端到端的多文档阅读理解模型 V-NET[29]。该模型主要利用基于注意力机制的多文档校验，解决歧义的挑战。V-NET 模型在 MSMARCO 数据集问答任务上三次获得第一[30]。

智能创作

百度大脑整合了自然语言处理和知识图谱技术，推出了智能创作平台[31]，提供自动写作和辅助创作能力，旨在成为最懂创作者的智能助手，全面提升内容创作效率和质量。在自动写作中，基于知识图谱的优质话题与素材库自动挖掘，为自动内容生成提供了优质候选；同时基于知识图谱，对话题人物、事件等素材进行有效、条理的组织，解决了传统基于检索等方式素材组织零散、逻辑性不强等问题。在辅助创作中，基于事件图谱的热点分析，可以快速发现热点，通过将事件脉络呈现给作者，激发作者写作灵感。知识图谱在话题挖掘、素材整理、热点发现，事件关联，事件脉络生成等智能创作核心技术中都起到重要作用。

在语言生成方面，百度研发了基于宏观规划、微观规划、表层实现的篇章生成算法，同时提出基于规划、信息选择、层次化等多种创新神经网络生成算法，在数据到文本生成、摘要生成、诗歌生成等任务上取得良好效果。目前，百度已将语言生成技术应用于百家号内容创作、语音播报、智能写作等，通过提供自动写作和辅助写作能力，提升内容创作的效率和质量，为智能创作领域提供更多可能。

五、平台层

自 2016 年以来，百度向广大开发者和企业用户开放 AI 能力。至 2019 年 7 月，百度已开放了 200 余项 AI 技术，并提供了一系列定制化 AI 开发平台，不断降低各行各业 AI 的应用门槛，助力各行业智能化升级。目前百度 AI 开放平台已形成了业界最全面和最领先的 AI 开放能力集合，以及最完备的 AI 开发平台矩阵，使用开发者已突破 150 万，并在持续快速增长，是业内服务规模最大的 AI 开放平台。

图 14 百度 AI 开放平台

作为中国服务规模最大的 AI 开放平台，已在语音、听觉、语言等方面实现全面升级，从深度学习框架、场景化 AI 能力、定制化训练平台，到软硬一体模组和解决方案等多方面，加速 AI 在工业、交通和医疗等领域的落地。

图 15 百度大脑开放能力不断丰富

5.1 飞桨（PaddlePaddle）深度学习平台

百度早在 2012 年就尝试将深度学习技术应用于语音识别、OCR 等业务，2013 年开始研发深度学习平台飞桨并于 2016 年开源，且大规模推广应用。2017 年，百度牵头成立深度学习技术及应用国家工程实验室[32]。如图 16 所示，百度在多年深度学习和人工智能技术的积累和业务实践的基础上，研发了“飞桨产业级深度学习开源开放平台”，并与华为等合作伙伴进行了软硬一体的合作优化。飞桨平台是核心技术领先，集深度学习训练和预测框架、模型库、开发套件、工具组件和服务平台等为一体，功能完备、全面开源开放的产业级深度学习平台。

图 16 飞桨产业级深度学习开源开放平台全景图

5.1.1 领先技术

开发便捷的产业级深度学习框架

飞桨深度学习框架采用基于编程逻辑的组网范式，对于普通开发者而言更容易上手，符合他们的开发习惯。同时支持声明式和命令式编程，兼具开发的灵活性和高性能。网络结构自动设计，模型效果超越人类专家。

支持超大规模深度学习模型的训练

飞桨突破了超大规模深度学习模型训练技术，实现了世界首个支持千亿特征、万亿参数、数百节点的开源大规模训练平台，攻克了超大规模深度学习模型的在线学习难题，实现了万亿规模参数模型的实时更新。

多端多平台部署的高性能推理引擎

飞桨不仅兼容其他开源框架训练的模型，还可以轻松地部署到不同架构的平台设备上。同时，飞桨的推理速度也是全面领先的。尤其经过了跟华为麒麟 NPU 的软硬一体优化，使得飞桨在 NPU 上的推理速度进一步突破。

面向产业应用，开源开放覆盖多领域的工业级模型库

飞桨官方支持 100 多个经过产业实践长期打磨的主流模型，其中包括在国际竞赛中夺得冠军的模型；同时开源开放 200 多个预训练模型，助力快速的产业应用。

5.1.2 工具组件

灵活、高效、易用是飞桨成功推广的重要原因。2019 年，飞桨全新发布或升级了包括 AutoDL 3.0、PaddleHub、PARL、PGL、VisualDL 等在内的一系列工具组件。

AutoDL：自动化深度学习

为了降低成本，提高效率，并降低对数据的依赖，减少人工干预，百度开发了 AutoDL 自动化深度学习建模工具。AutoDL 具备 AutoDL Design、AutoDL Transfer 和 AutoDL Edge 三种功能。2019 年，AutoDL 升级到 3.0 并通过飞桨正式开源。AutoDL 3.0 从生成模型的设计、

集成迁移学习和模型压缩适配三方面全面升级，基于强化学习的网络结构搜索算法，提供案例供开发者理解算法和快速验证。

AutoDL Design 能够根据用户提供的数据集，设计全新的深度学习模型，能够利用深度强化学习、差异化架构搜索和集成学习能力，为不同的任务构建更好的神经架构[33]。通过 AutoDL Design 技术设计的图像分类网络在 CIFAR10 数据集上正确率达到 98%，效果超过人类专家，业内领先。表 3 列举了在 CIFAR10 上用于图像分类的 AutoDL 和其他方法的效果比较，其中每种方法都只显示了最好精度。

表 3 在 CIFAR10 上用于图像分类的 AutoDL 和其他方法的效果比较

模型	准确率 Acc	模型	准确率 Acc
vgg_15_BN_64	93	resnet_v2_bottleneck_164	94.2
vgg_16	93.6	wide_resnet	95
resnet_32	92.5	densenet_BC_110_12	95.5
resnet_44	93	resnext_29_8x64d	96.2
resnet_56	93.3	shake_shake_64d_cutout	97.1
resnet_110	93.5	AutoDL 2.0	98.0

AutoDL Transfer 支持小数据建模，可以将预先训练的模型迁移到新的学习任务中，以提供快速的建模体验和增强的模型精度。在 AutoDL Transfer 中，百度提出了一种新颖的规范化的转移学习框架 DELTA，即 Deep Learning Transfer using Feature Map with Attention[34]，DELTA 通过对齐两个任务中两个网络的特征图来进行知识迁移。与经典方法相比，DELTA 拥有更高的精度。

AutoDL Edge 提供了不同的压缩方法，在保持建模精度的同时压缩深度学习模型[35]。它可以将深度学习模型部署到拥有不同算力、内存资源的硬件上，满足不同的能源消耗、响应时间需求。AutoDL Edge 利用参数共享的方法（简称 FSNet，Filter Summary Network），促使卷积网络中的卷积核共享参数从而实现对卷积网络大小的压缩。

PaddleHub：迁移学习

PaddleHub 拥有命令行工具和 Fine-tune API 两项功能：通过命令行接口，用户可以便捷获取飞桨生态下的预训练模型，无需编写代码，即可完成预训练模型预测；用户也可借助 PaddleHub Fine-tune API，使用少量代码就能完成迁移学习。

PARL（PAddlepaddle Reinfocement Learning）：强化学习

PaddlePaddle PARL 是百度推出基于飞桨的深度强化学习框架，具有可复制、可重用以及可扩展的特点。该框架具有高灵活性，能够支持可定制的并行扩展，覆盖 DQN、DDPG、PPO、IMPALA、A2C、GA3C 等主流强化学习算法。

2018 年，在机器人控制会议 CoRL 上，百度发表了干预强化学习机制的工作；同年在 NeurIPS 2018 举办的 AI 假肢挑战赛中，首次参赛就一举击败全球 400 多个参赛团队，以 9980 分的成绩夺得冠军，领先第二名高达 30 多分[36]。

PGL（Paddle Gragh Learning）：图神经网络

PGL 提供基于游走（Walk Based）以及消息传递（Message Passing）的两种计算范式，能够涵盖大部分前沿图学习算法，如 GCN、GAT、GraphSAGE、node2vec 等。PGL 充分利用 Paddle LoD Tensor 特性，大幅提升了消息传递范式中的信息聚合效率，兼顾了灵活性和高效性。PGL 利用图采样算子单机可支持 5 千万节点、20 亿条边的巨图。

VisualDL：训练可视化工具

VisualDL 能够帮助开发者方便地观测训练整体趋势、数据样本质量、数据中间结果、参数分布和变化趋势，以及模型的结构，更便捷地处理深度学习任务。

5.1.3 服务平台

AI Studio：AI 开发实训平台

AI Studio 集开放数据、开源算法、免费算力于一体，为开发者提供高效易用的学习和开发环境、丰富的体系化课程、海量开源实践项目、以及高价值的竞赛，并提供教育版支撑高校和机构老师轻松实现 AI 教学，助力深度学习人才培养。目前 AI Studio 平台上已开放数千个样例工程和数据集、数百节精品课程，并提供海量免费 GPU 算力资源，帮助开发者

训练了超过 10 万个实验项目。

EasyDL：定制化训练和服务平台

EasyDL 为企业用户和开发者提供高精度 AI 模型定制开发的全流程支持，包括数据标注与管理、模型训练、服务部署等各个环节完善的功能支撑，并与 EasyEdge 无缝对接实现定制模型的端侧部署。

EasyDL 面向零算法基础用户提供可视化操作界面，并支持自动化模型训练，大幅降低了 AI 开发与应用的门槛；面向专业算法工程师提供可灵活调参的开发界面，并预置丰富的优秀预训练模型和网络，帮助开发者高效获取高准确率模型；面向零售行业客户提供 EasyDL 零售版，针对商品检测场景专项调优算法，并提供商品采集箱、数据增强、货架拼接等配套增值服务，助力零售场景 AI 应用落地。

目前，EasyDL 支持图像分类、物体检测、图像分割、视频分类、声音分类、文本分类与匹配等多种模型类型，已有超过 2 万个用户在 EasyDL 上训练近 6 万个模型，并在零售、工业、安防、医疗、互联网、物流等 20 多个行业中落地应用。

EasyEdge：端计算模型生成平台

EasyEdge 是百度基于飞桨打造的零代码生成高性能端计算模型平台，为有边缘计算需求的开发人员提供自动模型转换、压缩、推理增强和 SDK 打包功能服务。平台目前支持飞桨、Caffe、PyTorch 和 TensorFlow 四种主流深度学习框架，13 种经典网络结构及部分自定义网络结构。

开发者不需要具备深厚的硬件开发能力，最快仅需 2 分钟即可生成端计算模型及封装 SDK。目前，EasyEdge 已针对通用 ARM 芯片、通用 x86 芯片、NVIDIA GPU、华为 HiSilicon NPU、英特尔 Movidius VPU、高通 Snapdragon GPU/DSP、苹果 A-Bionic 等多种芯片加速，生成适配的最优端计算模型。同时，EasyEdge 也与 EasyDL 定制化模型训练和服务平台进行了无缝对接，帮助用户轻松实现定制化模型的设备端计算和部署。

5.2 UNIT 智能对话训练与服务平台

UNIT（UNderstanding and Interaction Technology）为开发人员创建自己的对话系统提供了一个高性价比的解决方案。继 2018 年发布 UNIT 2.0、升级了先进的企业级对话系统功能

之后，2019 年 5 月 UNIT 升级到 3.0，在搭建技能、构建知识和整合技能与知识三方面实现了全面升级，并发布 9 大核心特性。UNIT 平台核心技术包括语义理解、阅读理解和对话管理三大部分，其中 ERNIE SLU 可达到在同样理解精度下标注量降低 37%至 72%，数据辅助生产工具 DataKit 可使数据生产效率提升 8 倍，使用语义理解 SLU 定制可使对话技能综合研发成本降低 60%。目前，UNIT 平台已实现 7.4 万定制技能，累计交互次数超 1380 亿次，全面覆盖智能客服、智能出行、智能办公及其他智能交互场景，为一线开发者实现 AI 产业化提供有力工具。

5.3 开放数据集

大规模领域数据集是机器学习研究和应用的基础。为了更好地促进算法研究，百度 BROAD—Baidu Research Open Access Dataset[37]项目发布了多个行业级的领域数据集，如大规模自动驾驶数据集（ApolloScape）、街景图像数据和文字场景数据集（ICDAR2019-LSVT，ICDAR2019-ArT）、视频精彩片段（标签视频序列和剪辑数据集）、场景解析（车载传感器捕捉的室外场景数据集）、机器阅读理解（DuReader）、中文句子及相应的开放域信息提取事实数据集（SAOKE）、中文信息抽取数据集（Schema based Knowledge Extraction，SKE）、中文短文本实体标注数据集（Baidu Entity Recognition and Linking Dataset，BERL）、中英同传数据集 BSTC（Baidu Speech Translation Corpus）、青光眼图像数据集（iChallenge-GON）、年龄相关性黄斑变性图像数据集（iCHallenge-AMD）、病理性近视（iChallenge-PM）等。

5.3.1 自动驾驶数据集 ApolloScape

作为首个来自大型自动驾驶公司的大规模开放数据集，ApolloScape 是行业内环境最复杂、标注最精准的三维自动驾驶公开数据集之一[38]。目前 ApolloScape 已经发布了 7 个公开数据集，涵盖了二维场景解析（Scene Parsing）、车道线细粒度检测、2D/3D 车辆自定位（Self-localization）、基于单帧图像的三维车辆位姿估计、车辆轨迹预测、基于点云的三维物体检测和跟踪以及双目深度估计等。

自 2018 年 3 月的发布以来[39]，ApolloScape 数据集已经在全球范围内被下载上万次。由于其重要性，计算机视觉领域内顶级的杂志 IEEE TPAMI 收录了关于这个数据集的科技文

章，这也是目前我们所知的唯一一篇发表在 TPAMI 上的关于自动驾驶数据集的文章。基于这些数据集，百度在 CVPR、ECCV 等世界级的顶会和 Kaggle 平台上举办了多项竞赛。这些竞赛吸引了来自全球范围内上千个参赛队伍，其中包括斯坦福大学、英伟达、旷世、优图等大学和企业。

图 17 ApolloScape 数据集示例：彩色图像（上部）及其二位语义标签（下部）

5.3.2 语言数据

在自然语言处理领域，百度发布了多个中文数据集，面向业界和学界，推动中文自然语言处理技术的发展。

阅读理解数据集 DuReader：百度发布的最大的面向搜索场景的阅读理解数据集 DuReader 2.0，包含了用户发布到百度搜索引擎的真实匿名查询中的 30 万个问题，150 万个来自网络的完整文档和 66 万个人工生成的答案。基于该数据，百度、中国计算机学会和中文信息学会连续两年举办了机器阅读理解评测，推动了中文阅读理解技术的进步。

信息抽取数据集 SKE：当前，阻碍中文信息抽取技术发展的瓶颈是缺乏大规模、高质量中文信息抽取语料。为了解决该问题，根据百度百科及信息流文本，百度构建并发布了业界规模最大的中文信息抽取数据集 SKE，包括 50 类关系，21 万中文句子及 43 万三元组实例，具有规模大、精度高、实战性强的特点。为了推动信息抽取技术的发展，百度联合中国计算机学会、中国中文信息学会举办“2019 语言与智能技术竞赛”信息抽取评测。评

测吸引了 1836 支队伍报名参加，在线评测提交结果 3300 余次，在业界和学界产生了广泛影响，推进了中文信息抽取技术的发展。

实体链指数据集 BERL：实体链指（Entity Linking）是 NLP 领域的基础任务之一。传统的实体链指任务主要是针对长文档，相比之下，短文本实体标注由于上下文语境不丰富、口语化等问题仍存在很大的挑战。为促进知识图谱领域的技术发展，百度发布了业界最大规模的中文短文本实体标注数据集 BERL，并联合中国中文信息学会语言与知识计算专委会共同举办了面向中文短文本的实体链指评测任务[40]。

机器同传数据集 BSTC：为了推动机器同传的相关研究和发展，百度发布了面向真实场景的中英同传数据集 BSTC。BSTC V1.0 发布了 50 小时的演讲数据，包含演讲者语音、文字转写、翻译结果等。基于该数据集，百度联合全国机器翻译大会 CCMT 举办了首届语音翻译评测。

5.3.3 医学影像数据

2018 年，百度与临床合作伙伴中山大学中山眼科中心联合发布 iChallenge 平台[41]，该平台旨在持续向公众分享已精细标注的眼部图像数据集，以供研究之用。迄今为止，iChallenge 分别发布了青光眼（iChallenge-GON）、病理性近视（iChallenge-PM）、闭角型青光眼（iChallenge-PACG）和黄斑变性（iChallenge-AMD）四个数据集，并分别在医学影像顶会 MICCAI 和 ISBI 上举办对应的现场比赛 REFUGE、PALM、AGE 和 ADAM。

六、AI 安全

人工智能商业化应用进入井喷阶段，相关的法规和标准、安全与隐私技术等面临着巨大的挑战。为此，百度组建了由顶尖安全专家构成的 AI 安全研究团队，在 AI 系统安全、数据安全、模型攻防等方向构建国际领先的关键支撑性技术，并推动生态合作与标准建设，与合作伙伴一起打造一个空前繁荣的领先安全技术生态。

AI 系统安全

百度推出了BASS OS（百度AI安全OS栈），为通用AIoT OS（Android/Linux/FreeRTOS）提供了“智能反应装甲”，可以为产业链中因为碎片化导致高危漏洞频出的通用嵌入式OS提供全栈的安全防护。其中 KARMA 自适应热修复技术可以跨系统版本与编译配置进行自适应漏洞修复。该技术能快速、全面、实时修复各种终端中的系统漏洞，使得漏洞修复不再依赖于冗长、繁杂的系统更新流程，直达终端用户。利用 KARMA 技术修复漏洞，无需依赖被修复系统的源码，且一个修复方案可以支持多种不同设备，极大减轻了终端厂商的漏洞修复成本，极大缩短了智能终端厂商推送系统安全补丁到最终用户的过程，对AI时代的智能终端的安全防护起到了至关重要的作用。自 2017 年 11 月成立的国内首个智能终端安全生态联盟 OASES，到现在汇聚了近 60 家不同的厂商和机构，包括海思、Mstar、全志等芯片企业，华为、中兴、酷派、长虹、TCL、暴风、康佳、蓝港、极米、联想、海尔等设备厂商，盘古、KEEN、NewSky Security、安天等安全厂商，以及清华、复旦、中科院、上海交大、弗罗里达州立大学等高校研究机构。

百度的形式化验证技术DeepVerification也是在国际上第一次为Apollo无人驾驶系统的核心模块这样的真实复杂系统进行安全性证明。同时DeepVerification对百度核心的开源项目如Apollo、飞桨、OTA，以及它们所依赖的第三方库超过10万行代码进行了验证和扫描，发现了超过70个高危安全漏洞，并做了相关通报。百度受邀在KLEE Workshop 2018上展示了该技术成果[42]，同时以该技术为主的学术论文将发表于IEEE Security & Privacy 2020 [43]。

数据安全与隐私保护

可信计算技术主要用于数据的安全流通以及AI模型保护，在不泄露隐私数据前提下，支持数据分析和机器学习等任务。2018 年百度联手 Intel 发布了全球首个内存安全的可信安全计算服务框架 MesaTEE（泛在安全计算技术），引领国际范围下一代数据安全技术的革命。通过结合 TEE 可信隔离保护和领先的 HMS 混杂内存安全模型，为 AI 生产力革命奠定了下一代生产关系的基础，让 AI 商业模式的命脉——安全的大数据供应链成为可能。MesaTEE与飞桨Anakin、TVM、GBDT等机器学习框架进行了紧密集成，为先进机器学习算法提供了顶级的数据与模型保护能力。

百度还开发了基于软件的安全多方计算方案，包括用于可信数据分析的 SQL 查询引擎和用于可信机器学习的逻辑回归联合建模实现。百度开发了一系列 MESA（Memory Safe）

为主的开源项目，包括 Rust SGX SDK（Interl SGX 可信计算平台的 Rust 语言开发包）、MesaLink（下一代传输层安全库）、MesaLock Linux（提供安全用户态环境的内存安全操作系统）、MesaPy（安全且快速的 Python）、MesaTEE GBDT-RS（机器学习算法 XGBOOST 在 SGX 中实现）。其中，MesaTEE 已经开源成为 Apache 软件基金会孵化项目。

此外，百度着力推动 MesaTEE 项目独创的 Hybrid Memory Safety（HMS）技术及理念，构造完整的内存安全系统，并在其上利用 Non-bypassable Security Paradigm（NbSP）等可扩展形式化验证技术保障整体应用的安全性，以构建坚实的防御阵地。

AI 模型攻防

在安全攸关场景中，AI 模型安全性不足容易造成严重威胁。例如在正常环境的概率扰动下，只要少许的扰动即可对模型造成100%的绕过；同样在数字空间白盒情况下，欺骗模型所需的扰动可以是非常细微的。在物理世界里，也可以通过贴片式攻击，对目标检测模型造成漏检。BlackHat Europe 2018 大会和 GeekPwn 2018 Las Vegas 大会上，百度成功实现了视觉感知模型的真实物理欺骗[44]，可以让物体“凭空消失”或者“无中生有”，证明了深度学习模型面临的严峻安全挑战。而对于云端黑盒模型，百度只需要极少的访问数量即可造成黑盒模型的绕过，从而实施对主流公有云平台的 AI 图像分类服务以及目标检测服务的欺骗。这项工作在 Blackhat Asia 2019[45]，Defcon China 2019[46]大会上成功演示。

基于这些工作，百度开源了 AI 对抗攻防工具包 Advbox，其中不但包括各种针对深度学习图像分类模型的基于白盒和黑盒的攻击算法；而且新增了对模型鲁棒性的标准化测量的工具 percetron robustness benchmark，为生产线上的模型鲁棒性评估提供了支撑。该工具支持飞桨、Tensorflow、Keras、PyTorch 等多平台，支持包括图像分类及目标检测等任务，支持标准化的度量，支持云端黑盒模型的鲁棒性度量，包括 Google Cloud Vision，Baidu AIP，Amazon Rekognition 等，使用形式化验证的方法，模型能够确保预测结果的一致性。百度在深度学习模型鲁棒性的研究中，率先提供了适应产业生态的模型可靠性评测，不但提供了超前的威胁感知能力，还提供了业界领先的对抗实施方案及实现。

大数据威胁情报

百度的反黑产安全 SDK 提供了移动端设备的威胁感知，包括设备环境威胁，代码逆向威胁和网络威胁，成为百度核心 App 的标配。在端上为百度庞大的移动生态提供账号安全、金融安全、数据资产安全等方面的贴身保护。在面对移动端强对抗的环境下，安全 SDK 依然能够保持对设备的唯一识别的能力以及安全风控因子的搜集。这些研究成果发表在 ICSE 2018[47][48]。同时该能力也赋予百度复合风控平台有效关联和识别与解决黄牛党、羊毛党进行刷单刷水、抽奖作弊、恶意下单以及移动 APP 渠道推广反作弊等问题。目前已帮助内外部数十家产品线提供高效稳定的活动防刷服务。

百度复合风控平台，基于长期一线与黑产多维度对抗积累的经验和庞大的移动端基数，借助百度独创的 5 层复合机器学习和威胁情报大脑双擎驱动，为各类线上活动提供了专业的反作弊能力，精准实现对规模化深层黑产感知、人机识别、群控加代理集群拦截及黑产溯源等功能。复合风控平台核心能力是智能“风控大脑”，基于百度强大的 AI 能力，针对不同的业务场景构建了复杂的模型系统，其主要思想是自动化处理各种业务场景下的数据，通过数据维度切分，特征提取，人工安全专家指导和无监督学习相结合的方式快速将黑产和正常用户区分出来，并进一步针对隐藏很深的黑产采用多种复杂的模型进行分析。比如，图智能分析方面，百度提出了基于黑产设备社交网络的风险检测算法[49]并在安全顶级会议 CCS 的 AI Security 分论坛做了报告，代表了百度 AI 技术在业务风控领域的领先性。此外，2019 年的春晚，百度春晚红包项目也由风控系统保驾护航，拦截风险 1.9 亿次，完美保障了春晚活动。

百度还推出了 HugeGraph 的开源项目，在反欺诈、知识图谱、机器学习等诸多领域有着重要的应用。它是解决反欺诈、威胁情报、黑产打击等业务的图数据存储和图建模分析需求，并在此基础上逐步扩展并支持了更多的通用图应用。

应用安全

百度以开源形式推出了 OpenRASP（运行时自我保护技术），是企业 Web 安全防护中的一个重要武器，有效增强防御体系纵深和对漏洞防护的适应能力。它实现了这样的安全框架：不依赖请求特征检测攻击，而是在应用执行上述关键操作时，执行一段自定义的逻辑检查是否存在异常。该技术的优势在于可以对抗未知漏洞，给出漏洞详情，极大降低误

报率，让应用从容应对未知风险。目前，OpenRASP 已经帮数十家企业修补百余个漏洞，并保障了 Apollo 无人驾驶系统的云端安全。

除了以上的研究成果之外，还在减少操作系统内核攻击面、SGX 侧信道攻击，内存 rowhammer 副作用产生的信息泄漏，特别是对芯片级别的 meltdown，spectr 及其衍生的漏洞产生的隐私问题有着世界领先的研究成果。百度多次获邀在 BlackHat Asia[50][51]，RAID 等安全会议[52]上演讲并交流相关工作。

结语与展望

人工智能技术正在高速发展，走进各行各业，改变着现有生产方式，提升企业生产效率。回首过去一年的发展，我们可以看到智能+产业解决方案已经日臻完善，人工智能正在推动我们进入标准化、自动化、模块化的工业大生产阶段。这将对经济发展持续产生巨大的影响，亦将重塑众多行业的竞争格局。

（一）AI 算法、算力技术持续突破，让机器初步拥有理解世界的能力，计算发生在设备（Device）、边缘（Edge）和云（Cloud）中，D-E-C 场景将会是关注重点

大规模符号神经网络将使机器具有一定的深度理解和推理能力。构建大规模图神经网络，结合符号系统和深度网络的优势，使得神经网络能结构化表示知识、具备理解和初步推理的能力，向通用人工智能（AGI）迈出重要的一步。

软件定义的芯片将逐步超越通用及半通用芯片，成为 AI 计算的重要力量。未来的芯片架构和产品定义，将和 AI 应用、算法和计算框架紧密结合，形成“算法-软件-硬件”三位一体协同设计的新型芯片。AI 芯片将进一步提高 AI 算法的性能和速度，让机器可以更准确、快速地认识和理解世界。

边缘计算正在向边缘 AI 模式转变，边缘计算解决方案将更加强调智能化。边缘计算被视为是实现物理计算、智能城市、普适计算和物联网的重要因素。AI 在边缘计算方面将会跟物联网融合，在公有云中运行的大多数模型将部署在边缘计算层面。

（二）AI 应用从点向面、从消费领域向产业领域加速辐射，AI 工业化大生产时代正在来临

以 AI 等新技术为核心的智能云将成为新的赛道，基于 AI 能力的云将深入渗透企业的业务层，推动行业和产业的转型升级。顺应产业智能化大潮，智能云将下沉行业场景，向垂直产业纵深发展，成为“智能+”的重要抓手和推动力量。

自动化深度学习建模工具规模化发展，适应不同商业场景。2019 年起，市场上将会出现更多的自动化深度学习软件包和更多基于自动化深度学习的应用，更多的云服务商将提供自动化深度学习能力，行业应用门槛不断降低。

5G 网络商用将会加速 AI 技术落地各垂直行业。5G 网络与 AI 技术结合将为各行各业带来新的互联互通机遇，助推科技发展到万物互联的新阶段。

自动驾驶与车路协同双轮驱动，融合发展，加速智能交通落地。车和路将更加智能，智能交通加速在园区、城市、高速等多样化场景中落地，有效提升通行效率，保障出行安全。

（三）人工智能伦规范和治理体系将逐步构建

人工智能伦理规范及相应的法规将逐步构建和完善。人工智能的可解释性、透明性、公平性、安全性等引起各方重视，其发展需符合人类伦理和价值观，并遵守相应的法律法规，以避免机器决策对现有人类伦理和社会秩序的冲击。

参考文献

[1] https://github.com/baidu-research/NCRF

[2] Dingcheng Li, Jingyuan Zhang, and Ping Li, "Representation Learning for Question Classification via Topic Sparse Autoencoder and Entity Embedding," IEEE Big Data 2018, Seattle, USA, 2018.

[3] Dingcheng Li, Jingyuan Zhang, Ping Li, "TMSA: A Mutual Learning Model for Topic Discovery and Word Embedding", in Proceedings of the 2019 SIAM International Conference on Data Mining.

[4] Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li, "Knowledge Graph Embedding Based Question Answering," ACM International Conference on Web Search and Data Mining (WSDM '19), Melbourne Australia, 2019.

[5] Dingcheng Li, Siamak Zamani, Ping Li and Jingyuan Zhang, "Integration of Knowledge Graph Embedding Into Topic Modeling with Hierarchical Dirichlet Process", Accepted to appear at NAACL 2019.

[6] Jiaji Huang, Qiang Qiu, Kenneth Church. "Hubless Nearest Neighbor Search for Bilingual Lexicon Induction". ACL 2019.

[7] Mingming Sun, Xu Li, Xin Wang, Miao Fan, Yue Feng, and Ping Li, "Logician: A Unified End-to-End Neural Approach for Open-Domain Information Extraction", in Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18), New York, USA, 2018.

[8] Mingming Sun, Xu Li, Ping Li, "Logician and Orator: Learning from the Duality between Language and Knowledge in Open Domain," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP '18), Brussels, Belgium, 2018.

[9] Xin Wang, Kun Fang, and Runyao Duan. "Semidefinite programming converse bounds for quantum communication". IEEE Transactions on Information Theory. Pages 2583-2592, Vol. 65, Issue: 4, April 2019.

[10] Kun Fang, Xin Wang, Marco Tomamichel, Runyao Duan, "Non-asymptotic entanglement distillation". IEEE Transactions on Information Theory. Pages 6454-6465, Vol. 65, Issue: 10, October 2019.

[11] Kun Fang, Xin Wang, Marco Tomamichel, Mario Berta, "Quantum Channel Simulation and the Channel's Smooth Max-Information". IEEE Transactions on Information Theory (in Press), September 2019.

[12] Xin Wang, Mark M. Wilde, Yuan Su, "Quantifying the magic of quantum channels". New Journal of Physics. Pages 103002, Vol. 21, Issue: 10, October 2019.

[13] Yuxuan Du, Tongliang Liu, Yinan Li, Runyao Duan, and Dacheng Tao. "Quantum Divide-and-Conquer Anchoring for Separable Non-negative Matrix Factorization". In the 27th International Joint Conference on Artificial Intelligence (IJCAI18). July 2018.

[14] Gilad Gour, David Jennings, Francesco Buscemi, Runyao Duan, and Iman Marvian. "Quantum majorization and a complete set of entropic conditions for quantum thermodynamics". Nature Communications. Vol. 9, Number 5352, December 2018.

[15] Wei Ping, Kainan Peng, and Jitong Chen. "ClariNet: Parallel wave generation in end-to-end text-to-speech". ICLR 2019.

[16] Kainan Peng, Wei Ping, Zhao Song, Kexin Zhao. "Parallel Neural Text-to-Speech". Submited to NeurIPS 2019.

[17] Ming Liu, Yukang Ding, Min Xia, Xiao Liu, Errui Ding, Wangmeng Zuo, Shilei Wen, STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing, CVPR 2019.

[18] Xu Tang, Daniel K. Du∗, Zeqiang He, and Jingtuo Liu. "PyramidBox: A Context-assisted Single ShotFace Detector". URL: https://arxiv.org/pdf/1803.07737.pdf

[19] Wei Li, Chengwei Pan, Rong Zhang, Jiaping Ren, Yuexin Ma, Jin Fang, Feilong Yan, Qichuan Geng, Xinyu Huang, Huajun Gong, Weiwei Xu, Guoping Wang, Dinesh Manocha, Ruigang Yang, "AADS: Augmented autonomous driving simulation using data-driven algorithms", Science Robotics, 2019, Vol. 4.

[20] Pingping Huang, Jianhui Huang, Yuqing Guo, Min Qiao and Yong Zhu. "Multi-grained Attention with Object-level Grounding for Visual Question Answering". ACL 2019.

[21] Wei He, Zhongjun He, Hua Wu and Haifeng Wang. "Improved Neural Machine Translation with SMT Features". In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), Pages 151-157.

[22] Daxiang Dong, Hua Wu, Wei He, Dianhai Yu and Haifeng Wang. "Multi-Task Learning for Multiple Language Translation". In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Pages 1723 - 1732.

[23] Hairong Liu, Mingbo Ma, Liang Huang, Hao Xiong and Zhongjun He. "Robust Neural Machine Translation with Joint Texual and Phonetic Embedding". ACL 2019.

[24] Mingbo Ma, Liang Huang, Hao Xiong, Renjie Zheng, Kaibo Liu, Baigong Zheng, Chuanqiang Zhang, Zhongjun He, Hairong Liu, Xing Li, Hua Wu and Haifeng Wang. "STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework". ACL 2019.

[25] Hao Xiong, Zhongjun He, Hua Wu and Haifeng Wang. "Modeling Coherence for Discourse Neural Machine Translation". AAAI-2019.

[26] Yuchen Liu, Hao Xiong, Zhongjun He, Jiajun Zhang, Hua Wu, Haifeng Wang and Chengqing Zong. "End-to-End Speech Translation with Knowledge Distillation". InterSpeech 2019.

[27] 10 Breakthrough Technologies 2018, MIT Technology Review. URL: https://www.technologyreview.com/lists/technologies/2018/

[28] An Yang, Quan Wang, Jing Liu, Kai Liu, Yajuan Lyu, Hua Wu, Qiaoqiao She, Sujian Li. "Enhancing Pre-Trained Language Representations with Rich Knowledge for Machine Reading Comprehension". ACL 2019.

[29] Yizhong Wang, Kai Liu, Jing Liu, Wei He, Yajuan Lyu, Hua Wu, Sujian Li, Haifeng Wang, "Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification". ACL 2018.

[30] http://www.donews.com/news/detail/4/2987861.html

[31] https://ai.baidu.com/creation/main/index

[32] https://www.dlnel.org/

[33] Yingzhen Yang, Xingjian Li, and Jun Huan. "An Empirical Study on Regularization of Deep Neural Networks by Local Rademacher Complexity". arXiv preprint arXiv:1902.00873, 2019.

[34] Xingjian Li, Haoyi Xiong, Hanchao Wang, Yuxuan Rao, Liping Liu, and Jun Huan. "DELTA: DEep Learning Transfer using Feature Map with Attention for Convolutional Networks". the Seventh International Conference on Learning Representations (ICLR'19).

[35] Yingzhen Yang, Nebojsa Jojic and Jun Huan. "FSNet: Compression of Deep Convolutional Neural Networks by Filter Summary". arXiv preprint arXiv:1902.03264, 2019.

[36] NeurIPS 2018: AI for Prosthetics Challenge. URL: https://www.crowdai.org/challenges/neurips-2018-ai-for-prosthetics-challenge/leaderboards?challenge_round_id=64

[37] http://ai.baidu.com/broad

[38] http://apolloscape.auto/

[39] Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, Ruigang Yang, "The ApolloScape Dataset for Autonomous Driving", CVPRW 2018.

[40] https://biendata.com/competition/ccks_2019_el/

[41] https://ichallenge.baidu.com

[42] Peng Li, Rundong Zhou, Yaohui Chen, Yulong Zhang, Tao Wei, "ConcFuzzer: a sanitizer guided hybrid fuzzing framework leveraging greybox fuzzing and concolic execution", KLEE Workshop 2018.

[43] Yaohui Chen, Peng Li, Jun Xu, Shengjian Guo, Rundong Zhou, Yulong Zhang, Tao Wei, Long Lu, "SAVIOR: Towards Bug-Driven Hybrid Testing", 40th IEEE Symposium on Security and Privacy, 2020.

[44] Zhenyu Zhong, Weilin Xu, Yunhan Jia, Tao Wei, "Perception Deception: Physical Adversarial Attack Challenges and Tactics for DNN-Based Object Detection", BlackHat Europe, 2018.

[45] Yunhan Jia, Zhenyu Zhong, Yulong Zhang, Tao Wei, Yantao Lu, "The Cost of Learning from the Best: How Prior Knowledge Weakens the Security of Deep Neural Networks", BlackHat Asia, 2019.

[46] Yna Liu, Xin Hao, Yang Wang, Tao Wei, "Transferability of Adversarial Examples to Attack Cloud-based Image Classifier Service", Defcon China 2019.

[47] Pei Wang, Qinkun Bao, Li Wang, Shuai Wang, Zhaofeng Chen, Tao Wei, Dinghao Wu, "Software Protection on the Go: A Large-Scale Empirical Study on Mobile App Obfuscation", International Conference on Software Engineering, 2018.

[48] Pei Wang, Dinghao Wu, Zhaofeng Chen, Tao Wei, "Protecting Million-User iOS Apps with Obfuscation: Motivations, Pitfalls, and Experience", International Conference on Software Engineering, 2018.

[49] Chao Xu, Zhentan Feng, Yizheng Chen, Minghua Wang, Tao Wei, "FeatNet: Large-scale Fraud Device Detection by Network Representation Learning with Rich Features", 11th ACM Workshop on Artificial Intelligence and Security, 2018.

[50] Yueqiang Cheng, Zhi Zhang, Surya Nepal, Zhi Wang, "Winter is Coming Back: Defeating the Most Advanced Rowhammer Defenses to Gain Root and Kernel Privileges", BlackHat Asia, 2019.

[51] Yueqiang Cheng, Zhaofeng Chen, Yulong Zhang, Yu Ding, Tao Wei, "Oh No! KPTI Defeated, Unauthorized Data Leakage is Still Possible", BlackHat Asia, 2019.

[52] Zhi Zhang, Yueqiang Cheng, Surya Nepal, Dongxi Liu, Qingni Shen, Fethi Rabhi, "KASR: A Reliable and Practical Approach to Attack Surface Reduction of Commodity OS Kernels", 21st International Symposium on Research in Attacks, Intrusions and Defenses, 2018.


Bai du 百度