

# 边缘智能中的协同计算技术研究

□文 / 张星洲、鲁思迪、施巍松



张星洲

中国科学院大学计算技术研究所读博士。研究方向是人工智能、边缘计算、计算机系统。博士期间的研究项目主要包括基于云-边缘协同的驾驶行为分析、非侵入式智能家电感知系统、边缘计算数据框架性能分析和设计等。

美国韦恩州立大学在读博士。研究方向是边缘计算和分布式系统。博士期间的研究项目主要包括基于深度学习的大规模时间序列数据的异常检测和预测、边缘设备端的分布式协同合作，以及计算机视觉（多摄像头多人物追踪、性别年龄预测等）。

鲁思迪



施巍松

美国韦恩州立大学工学院副院长、博士生导师、教授，IEEE（电气和电子工程师协会）会士，ACM（国际计算机协会）杰出科学家，计算机科学系 Charles H. Gershenson 杰出教授，移动与互联网系统结构实验室主任，智能驾驶实验室主任，校级信息物理融合系统 (CPS) 和智能健康计划项目主任。研究方向包括计算机系统、边缘计算、可持续计算、智能网联与自动驾驶。

边缘智能的发展中面临着三个矛盾：智能算法的资源需求与边缘设备受限于资源之间的矛盾、服务质量与隐私保护之间的矛盾、智能任务需求多样与边缘设备能力单一之间的矛盾。通过边缘与云端、物端设备之间的协同计算可以有效地解决这些矛盾。本文归纳了目前存在的四种协同模式，分别是：边云协同、边边协同、边物协同和云边物协同。本文针对每一种协同模式，介绍了具体的协同方式、相关技术和实现方法。随后，以典型的边缘智能场景（网联汽车和智慧家庭）为例，分析协同计算的优势。最后，本文提出为了实现边缘智能中的真正协同需要面对的几个挑战。

## 一、边缘智能产生的背景

边缘智能是在边缘计算技术和人工智能应用的双重作用力的推动下发展的。在边缘智能的发展中存在三个矛盾：智能算法的资源需求与边缘设备受限于资源之间的矛盾、服务质量与隐私保护之间的矛盾、智能任务需求多样与边缘设备能力单一之间的矛盾。而利用边缘与云端、物联网设备端的协同计算，可以有效解决这些问题。

### 边缘智能

近年来，随着万物互联时代的快速到来和无线网络的普及，网络边缘的设备数量和所产生的数据都在快速增长。根据 IDC 预测，到 2025 年，全球数据总量将大于 180 泽字节 (zettabyte, ZB)，而物联网所产生的数据的 70% 以上都将在网络边缘进行处理 [1]。在这种情形下，以云计算模型为核心的集中式处理模式将无法高效处理边缘设备产生的数据。传统云计算存在几点不足：实时性不够；带宽不足；能耗较大；不利于数据安全和隐私 [2]。为了解决以上问题，面向边缘设备所产生的海量数据来进行计算的边缘计算模型应运而生。边缘计算是在网络边缘执行计算的一种新型计算模型 [3][4]。边缘计算操作的对象包括来自于云服务的下行数据和来自于万物互联服务的上行数据。而边缘计算的“边缘”是指从数据源到云计算中心路径之间的任意计算和网络资源，是一个连续统。

边缘计算的发展历经技术储备期和快速增长期，如今正处于稳健发展期 [3]。相比于传统云计算，边缘计算具备几个明显的优点：首先，数据在边缘端做处理，极大地减轻了网络带宽和数据中心的压力。其次，服务请求不需要等待云计算中心的处理结果，减少了网络延迟，提升了服务质量。最后，用户隐私数据不再上传，而是存储在网络边缘设备上，减少了网络数据泄露的风险，保护了用户数据安全和隐私。

以深度神经网络为代表的深度学习算法在诸多人工智能相关的应用中取得了很好的效果，例如图片识别、语音识别等，本文将这类算法统称为“智能算法”。传统的人工智能应用采取云计算模式，即将数据传输至云端，然后在云端运行算法并返回结果。随着边缘计算的发展，越来越多的人工智能应用迁移到边缘进行。边缘计算和人工智能的结合孕育出了边缘智能。边缘智能来自于边缘计算的推动和智能应用的牵引的双重作用。在边缘计算方面，物联网数据、边缘设备、存储、无线通信和隐私安全技术的成熟共同推动了边缘智能的发展。同时，智慧医疗、智能网联汽车、智慧社区、智能家庭和公共安全等人工智能应用场景的发展也促使边缘智能进一步发展。

在 OpenEI: An Open Framework for Edge Intelligence (OpenEI: 一种边缘智能开源

[1] Zwolenski M, Weatherill L. The digital universe: Rich data and the increasing value of the internet of things[J]. Australian Journal of Telecommunications and the Digital Economy, 2014, 2(3): 47.

[2] 施巍松, 张星洲, 王一帆, 等. 边缘计算: 现状与展望[J]. 计算机研究与发展, 2019, 56(1): 69-89.

[3] Shi W, Cao J, Zhang Q, et al. Edge computing: Vision and challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646.

[4] 施巍松, 孙辉, 曹杰, 等. 边缘计算: 万物互联时代新型计算模型[J]. 计算机研究与发展, 2017: 1.

框架) [5] 一文中, 边缘智能被定义为使边缘设备能执行智能算法的能力。这个能力包含 4 个元素 (ALEM): 准确率 (Accuracy), 边缘智能算法的准确率是该能力可用性的表现; 延迟 (Latency), 延迟衡量了训练好的模型在边缘进行推理所需的运行时间, 强调了边缘应用的实时性需求; 功耗 (Energy), 指执行智能算法过程中边缘设备消耗的能量; 内存 (Memory Footprint), 指智能算法对边缘设备的内存占用量。

边缘智能的生态系统涵盖多个领域, 既包含智能硬件、机器学习算法库、嵌入式操作系统、智能算法等, 又包含上层应用的设计等, 同时也有许多工作研究边缘-边缘和边缘-云等协同/调度方式。边缘智能的发展对边缘计算和人工智能计算具有双向共赢的优势: 一方面, 边缘数据可以借助智能算法释放潜力, 提供更高的可用性。随着万物互联的发展, 边缘数据需要借鉴智能算法进行分析, 实现智能化的功能。以深度学习为代表的智能算法能从这些数据中提取实时信息以提高决策的效率和准确性。今后有很多的物联网应用将会加入深度学习模块, 前景巨大。另一方面, 边缘计算能为智能算法提供更多的数据和应用场景。深度学习的四大推动力是算法、硬件、数据和应用场景, 传统的数据都存放在数据中心。而现在, 边缘端将提供极大的数据量, 推动人工智能的发展。

### 边缘智能发展中的矛盾

边缘智能的发展非常迅速。然而, 将资源需求量较大的人工智能算法由云端迁移到边缘仍然存在很多挑战。本章总结出三个主要的矛盾:

#### 智能算法的资源需求与边缘设备受限于资源之间的矛盾

这个矛盾是目前边缘智能发展中的主要矛盾, 也是研究的热点问题。以深度

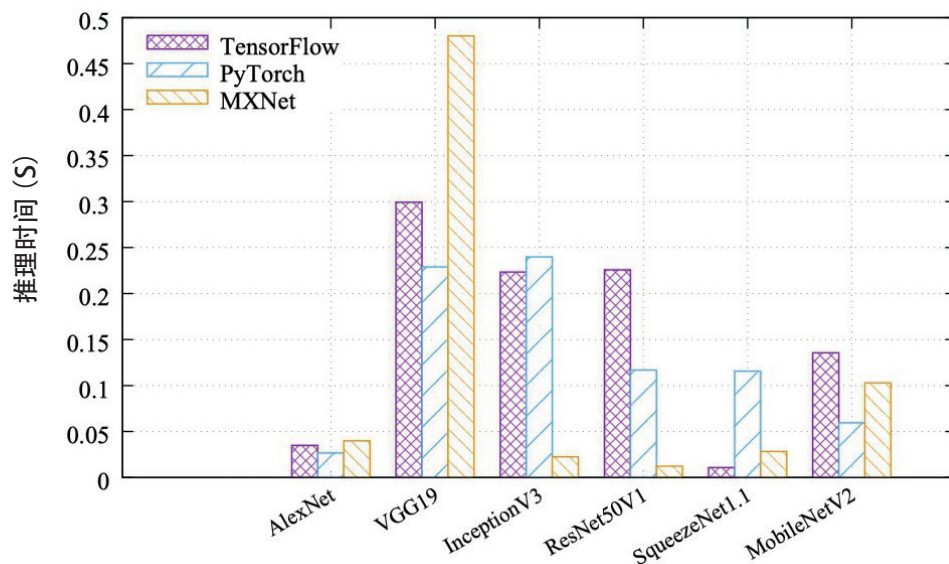


图 1: 图片分类算法在边缘设备上的运行时间



学习为代表的人工智能算法一般需要较强的算力支持，而与云端计算中心相比，边缘设备的计算能力相对受限，因此无法直接将云端算法迁移到边缘设备上执行。如图 1 所示，本文在 Intel 的边缘设备 FogNode 上运行 ImageNet 的图片分类任务，使用 TensorFlow、PyTorch 和 MXNet 等常见的软件作为计算框架，使用 AlexNet、VGG19、InceptionV3、ResNet50V1、SqueezeNet1.1 和 MobileNetV2 等算法作为计算负载。可以发现，对一帧图片的平均分类时间为 0.12s，最大时间为 0.48s。这无法满足边缘智能应用的实时性需求。同时，观察内存占用量，如图 2 所示，平均内存占用量为 700MB，最大可达内存占用量为 2000MB。内存资源受限的边缘设备无法承担，尤其是在多任务同时运行在边缘设备上时。

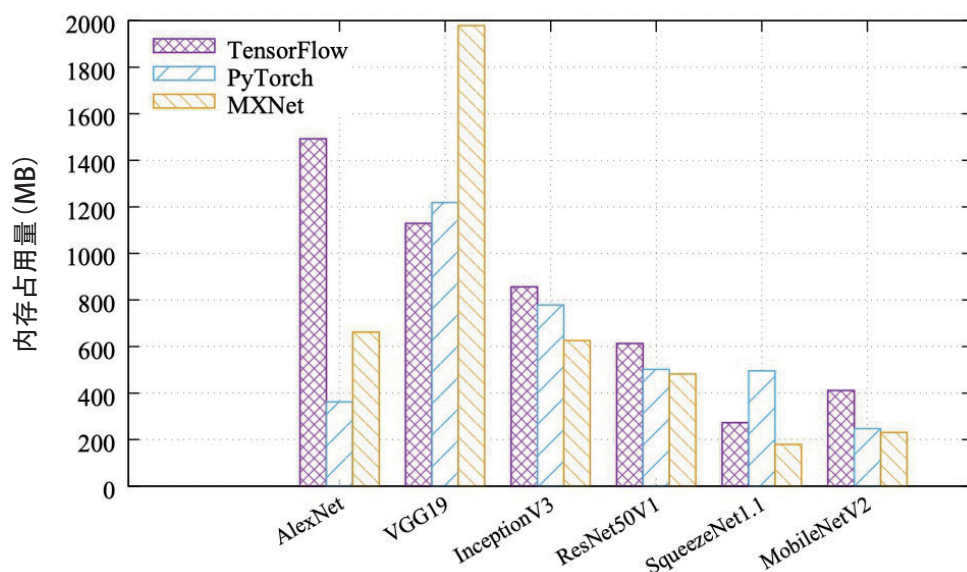


图 2: 图片分类算法在边缘设备上的内存占用量

### 服务质量与隐私保护之间的矛盾

边缘计算的一个优势是将数据保存在数据生产者的位置，这样可以最大限度地不暴露用户的隐私。但是人工智能算法的设计与提升往往需要大量且多样化的数据。算法得不到足够的数据进行训练，不利于智能算法的设计与提高，降低算法的准确率，从而最终影响服务质量。此外，数据在不同的边缘上相互独立存储，独立维护，彼此间相互孤立，容易形成数据孤岛。因此，在边缘智能场景下，存在服务质量与隐私保护之间的矛盾。

### 智能任务需求多样与边缘设备能力单一之间的矛盾

在边缘智能场景下，完成一项完整的智能任务往往需要经过多个阶段，具有多样化的需求，包括数据采集、预处理、计算、控制和反馈等。而由于地理空间等条件的限制，边缘设备只具有单项能力，例如只负责完成计算任务，这难以满足智能任务的需求。以

人脸识别开门的应用为例,人脸识别的任务可以划分成摄像头采集数据、边缘设备识别人脸、智能门锁受控打开等多项子任务。边缘端无法完成全部的子任务,需要与摄像头传感器、智能门锁等物联网设备进行协同。如果需要远程服务,则还需要云端的配合。

### 边缘智能中的协同计算技术

在边缘智能场景下应用协同计算技术可以有效地解决上述三个矛盾。

针对矛盾 1 (智能算法的资源需求与边缘设备受限于资源之间的矛盾),需要边缘与云端之间进行协同,合理优化任务分配策略,拆解智能算法,利用云端强大的计算能力承担公共的计算任务,减轻边缘的计算压力,满足智能算法、算力需求。此外,边缘与边缘之间的相互协作也可以进行减轻单一边缘节点的计算压力。

针对矛盾 2 (服务质量与隐私保护之间的矛盾),需要边缘与边缘之间进行协同,建立安全的通信机制,增强边缘设备之间的数据共享与协同能力,从而在保护数据隐私的情况下提高服务质量。

针对矛盾 3 (智能任务需求多样与边缘设备能力单一之间的矛盾),需要边缘与物联网设备之间进行协同,利用物联网设备多功能、空间分布多样化的优势,对任务进行协同处理,增强系统的整体能力。此外,边缘也需要与云端之间进行协同,增强对外服务能力,提升服务的多样性。

## 二、协同计算技术

在边缘智能场景下存在哪些协同模式?不同模式下的不同计算设备是如何协同分工的?为了实现这些协同,需要利用哪些技术,这些技术的成熟度如何?本章将边缘智能中的协同分成四类模式,分别是:边云协同、边边协同、边物协同和云边物协同。针对每一种模式,本章介绍了具体的协同方式、相关技术和实现方法。图 3 简述了本章的结构。

### 边云协同

#### 背景介绍

边缘与云端的协同是目前研究者和产业界探索最多的一种协同形式,也是技术阶段相对成熟的协同模式。在产业界,已经有多个组织/公司发布了基于边云协同概念的产品。2017年,工业互联网产业联盟(AII)在《工业互联网平台白皮书(2017)》中提出了工业互联网平台功能架构[6],初步呈现了边云协同。2018年,华为在其

[6] 工业互联网产业联盟. 工业互联网平台白皮书, 2018.

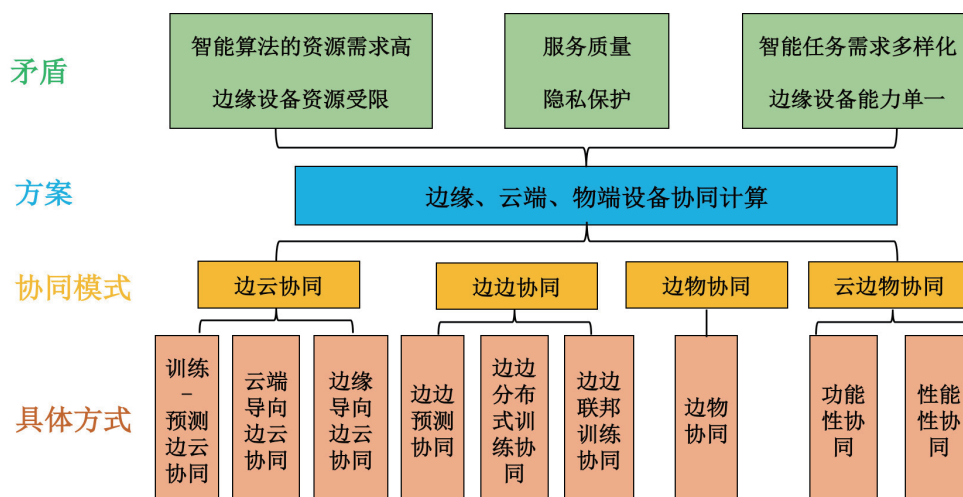


图 3：第二章结构介绍

Huawei Connect 2018 大会发布智能边缘平台 IEF ( Intelligent EdgeFabric )，其目标是满足客户对边缘计算资源的掌控管理和智能化的诉求，支持海量边缘节点安全接入、边缘应用生命周期管理，呈现出边云协同的一体化服务的理念。2018 年，KubeEdge 作为云边协同的开源智能边缘平台被提出，以支持云原生边缘计算 [7]。2019 年 7 月，云计算开源产业联盟发布了《云计算与边缘计算协同九大应用场景》白皮书 [8]。该白皮书认为，边缘计算是云计算的延伸，二者是相依而生、协同运作的。在边云协同中，边缘端负责在本地范围内的数据计算和存储工作，云端负责大数据的分析挖掘和算法训练升级。

### 协同方式

在边云协同计算中，云端和边缘有三种不同的协同计算方式：

**第一种：训练 - 预测边云协同 [9][10]。**这种情况下，云端根据边缘上传的数据来设计、训练智能模型并且不定期升级模型，边缘端负责搜集数据并且使用最新的模型预测实时数据。该协同方式比较成熟，已经应用于无人驾驶、视频检测等多个领域。谷歌公司推出的 TensorFlow Lite 框架即为该种类型的协同服务。它的运行方式是先在云端使用 TensorFlow 训练模型，然后下载到边缘端，TensorFlow Lite 加载模型，使用优化技术完成并加速预测任务。

**第二种：云导向的边云协同。**这种情况下，云端除了承担模型的训练工作之外，还会负责一部分预测工作。具体而言，神经网络模型将会被分割，云端承担模型前端的计算任务，然后将中间结果传输给边缘，边缘端继续执行预测工作，得出最终结果。该协

[7]Xiong Y, Sun Y, Xing L, et al. Extend Cloud to Edge with KubeEdge[C]//2018 IEEE/ACM Symposium on Edge Computing (SEC). IEEE, 2018: 373-377.

[8] 边缘计算产业联盟, 工业互联网产业联盟. 边缘计算与云计算协同白皮书 (2018 年), 2018,11.

[9]Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo and J. Zhang, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing," in Proceedings of the IEEE, vol. 107, no. 8, pp. 1738-1762, Aug. 2019.

[10]J. Chen and X. Ran, "Deep Learning With Edge Computing: A Review," in Proceedings of the IEEE, vol. 107, no. 8, pp. 1655-1674, Aug. 2019.

同方式的重点是找到合适的切割点，在计算量和通信量之间做权衡。该协同方式目前处于研究阶段，真实场景的应用还比较少。

**第三种：边缘导向的边云协同。**这种情况下，云端只负责初始的训练工作，模型训练完成之后下载到边缘上，边缘端除了实时预测之外，还会承担训练的任务，训练的数据来自于边缘自身的数据。这样得到的最终模型能够更好地利用数据的局部性，满足个性化的需求。该协同方式目前也处于研究阶段。

### 关键技术及相关研究

**迁移学习。**迁移学习技术在传统边云协同计算中的应用比较广泛 [11]。迁移学习的初衷是节省人工标注样本的时间，让模型可以通过已有的标记数据向未标记数据迁移，从而训练出适用于目标领域的模型。在边缘智能场景下往往需要将模型落地，即适用于不同的场景。以人脸识别应用为例，不同公司的人脸识别门禁一般使用相同的模型，然而训练模型原始的数据集与不同公司的目标数据集之间存在较大的差异。因此，可以利用迁移学习技术，保留模型的原始信息，然后加上新的训练集进行学习更新，从而得到适用于某一个边缘场景的模型。

**神经网络拆分。**在云导向的边云协同中，需要将神经网络模型进行拆分，一部分在云端执行，一部分在边缘端执行。因此，需要找到合适的切割点，尽量将计算复杂的工作留在云端，然后在通信量最少的地方进行切割，将中间结果传输至边缘，实现计算量和通信量之间的权衡。Kang、Hauswald、Gao 等人在 2017 年提出的 Neurosurgeon[12] 便是其中的代表性研究。它用一个基于回归的方法来估计 DNN（Deep Neural Networks，深度神经网络）模型中每一层的延迟，然后返回最优的分割点以达到延迟目标或能耗目标。

**神经网络压缩。**在边缘导向的边云协同中，边缘的计算能力相对受限。为了实现边缘端的有效训练，减少算力是关键需求。目前，许多针对神经网络的压缩技术在解决这一问题。一般而言，通过参数共享和裁剪等方式减少不敏感的参数，可以降低存储和通信开销，减少网络规模，降低算力需求。针对权值的量化等工作也可以减少网络的计算量 [13]。

## 边边协同

### 协同方式

边缘与边缘之间互相协同是目前的研究热点，其主要解决矛盾 1 和矛盾 2。具体而言，一是单个边缘的计算能力有限，需要多个边缘之间进行分时配合，提升系统整体的能力。例如，在完成神经网络模型的训练任务时，单个边缘进行训练比较吃力，既耗费较大时间和算力，又容易因为数据量的限制使得模型过拟合。因此，需要多个边缘协同训练。

[11]Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering, 2009, 22(10): 1345-1359.

[12]Kang Y, Hauswald J, Gao C, et al. Neurosurgeon: Collaborative intelligence between the cloud and mobile edge[C]//ACM SIGARCH Computer Architecture News. ACM, 2017, 45(1): 615-629.

[13]Han S, Mao H, Dally W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding[J]. arXiv preprint arXiv:1510.00149, 2015.

二是解决数据孤岛的问题。边缘的数据来源具有较强的局部性，需要与其他边缘协同以完成更大范围的任务。例如，在交通路况监测中，一般一个边缘只能获取当地的路况信息，多个边缘间的相互协作可以组合成大区域的路况地图。

边缘与边缘有三种协同方式：

**第一种：边边预测协同。**在这种方式下，云端先训练好模型，然后根据边缘的算力情况拆分模型并分配到边缘设备上。这样每个边缘设备执行一部分模型，可以减少计算压力。这种边缘协同方式一般适用于手机、手环等计算能力十分受限的边缘之间。2017年，一个为DNN模型设计的本地分布式移动计算系统MoDNN被提出[14]。它将一个已经训练好的模型拆分到多个移动设备上执行，多个移动设备通过无线连接建立一个小规模的计算集群。实验表明，当移动设备增加到2-4个时，MoDNN可以加速DNN执行时间2.17-4.28倍。

**第二种：边边分布式训练协同。**在这种协同方式下，每个边缘都作为计算节点承担人工智能模型的训练任务，边缘上拥有整个模型或者部分模型，训练集来自于边缘自身产生的数据。模型训练到一个阶段后，会将训练得到的模型参数更新到中心节点（参数服务器）中，最终训练得到完整模型。在这种协同中，设计高效的参数更新算法以达到带宽和模型准确率的权衡是研究的热点[15]。

**第三种：边边联邦训练协同。**联邦学习的协同是基于数据安全和隐私的目的提出的[16]。这种方式下，某个边缘节点保存最优模型，每个边缘作为计算节点参与模型的训练，其他节点在不违反隐私法规的情况下向该节点更新参数。该协同方式与边边分布式训练协同有相似之处，而它们的区别在于：在联邦训练协同中，边缘节点是数据的绝对拥有方，可以自主决定参与学习的时机；而在分布式训练中，中心节点占据主动地位，对分节点具有管理权限。在设计目标上，联邦训练协同中模型的更新更加侧重于数据的隐私，而分布式训练协同更加侧重于充分利用边缘节点的闲置资源。

## 关键技术及相关研究

**模型拆分。**与边云协同相似，在边缘之间协同也需要模型的分割，而且针对不同的边缘以及边缘资源的动态性，需要分割的次数和切割点的位置更加多样化。2018年，密歇根州立大学提出一种新型计算框架NestDNN[17]，将运行时资源的动态变化纳入考量，生成一种资源感知的深度学习移动视觉系统。NestDNN使得每个深度学习模型能够进行灵活的资源-准确率权衡。实验显示，NestDNN的推断准确率提高了4.2%，视频帧处理速率提高了1倍，能耗降低了40%。

**分布式训练。**分布式训练的思想在云计算中心的应用比较广泛[18]，也相对成熟。

[14]Mao J, Chen X, Nixon K W, et al. Modnn: Local distributed mobile computing system for deep neural network[C]//Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017. IEEE, 2017: 1396-1401.

[15]Wang S, Tuor T, Salonidis T, et al. When edge meets learning: Adaptive control for resource-constrained distributed machine learning[C]//IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018: 63-71.

[16]McMahan B, Ramage D. Federated learning: Collaborative machine learning without centralized training data[J]. Google Research Blog, 2017, 3.

[17]Fang B, Zeng X, Zhang M. Nestdnn: Resource-aware multi-tenant on-device deep learning for continuous mobile vision[C]//Proceedings of the 24th Annual International Conference on Mobile Computing and Networking. ACM, 2018: 115-127.

[18]Dean J, Corrado G, Monga R, et al. Large scale distributed deep networks[C]//Advances in neural information processing systems. 2012: 1223-1231.



在边缘智能场景下,该分布式训练的意义仍然存在,可以利用边缘的闲置资源进行模型的训练工作。但在边边协同的情况下,存在两个很大的区别:首先,云计算中心的节点计算能力大体相似,而边缘节点的计算能力差别极大,并且由于边缘也要承担其它的计算任务,同一个边缘节点在不同时刻的计算能力也有差异。因此,需要考虑到边缘设备资源的动态变化性。其次,云计算中心节点间采用以太网甚至光纤直接连接,而边缘节点由于地理空间的限制,它们之间的连接需要经过多次跳转,速度和带宽相对较差。因此,在设计分布式训练策略时也要考量通信质量。

**联邦学习。**2017年,谷歌利用联邦学习解决移动设备在本地更新模型的问题[16]。之后,联邦学习被推广至医疗、金融等领域。由于这些场景下的数据更加敏感和隐私,联邦学习提出在保障数据交换时安全和隐私的前提下,利用多个计算节点进行模型更新。其具体做法是在一个公共节点上建立一个虚拟的共有模型,其他节点在不违反隐私法规的情况下向该节点更新参数,最终的结果是该虚拟模型将全部的数据聚合在一起形成最优模型。2019年,谷歌推出了支持联邦学习的计算框架[19]。

## 边物协同

### 协同方式

边物协同中的“物”指物端设备,也可被表述为物联网设备,主要包括传感器设备、摄像头、工厂机械设备等。边物协同主要解决矛盾3,即增强边缘节点的能力。该协同在物联网,尤其是在智能家居和工业物联网中的应用非常广泛。边物协同下,物端负责采集数据并发送至边缘,同时接收边缘的指令进行具体的操作执行;边缘负责多路数据的集中计算,发出指令,对外提供服务。由于物联网设备与用户的结合更加紧密,因此,边物协同被认为是人工智能应用落地的关键一环。

### 关键技术及相关研究

**轻量级模型。**在边物协同下,边缘作为计算任务的主体和系统的核心中枢,需要承担更多的计算任务。因此,对于模型的压缩和设计轻量级模型是其中的关键技术。研究人员针对这种资源受限场景下的计算设备设计了轻量级的神经网络模型。2016年,SqueezeNet[20]作为小型的DNN架构被提出。它在ImageNet上实现了AlexNet的精度,但是其参数数量却是原来的五分之一。谷歌为移动视觉应用提供了有效的CNN模型MobileNets[21],其引入的两个超参数允许模型构建器为特定应用程序选择合适大小的模型。MobileNets主要由深度可分离卷积生成,这些卷积首先在图片分类的工作中引入,随后用于目标识别等。

**模型选择。**对于深度学习模型而言有一个不严谨的趋势——模型的规模越大(参

[19]Bonawitz K, Eichner H, Grieskamp W, et al. Towards federated learning at scale: System design[J]. arXiv preprint arXiv:1902.01046, 2019.

[20]Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[21]Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

数的数量越多)，其准确率越高。然而，在边缘智能计算场景下，一方面，资源的限制无法承担大规模的模型；另一方面，场景的不同也对准确率有不同的要求。在很多情况下，可以牺牲一部分准确率以换取较好的实时性。模型选择技术便是在模型的资源耗费量和准确率之间寻求较好的权衡，得到最符合场景需求的模型。

### 云边物协同

#### 协同方式

云边物协同为将整个链路上的计算资源都充分利用，发挥不同设备的优势，全方位地解决以上三个矛盾。云边物的协同分为两种协同方式：一是功能性协同。这种协同基于不同设备所处的地理空间、所担当的角色等不同而承担不同的功能。例如，由物端负责采集、边缘负责预处理、云端负责多路数据的处理和服务提供等。二是性能性协同。这是由于算力的限制，不同层级的计算设备承担不同算力需求的任务，包括任务的纵向切割和分配等。

#### 关键技术

云边物协同所利用的技术涵盖了前文介绍的各类技术，包括轻量级模型的设计、模型的分割和选择、分布式训练、联邦学习和迁移学习等，在不同的场景下有不同的应用。除此之外，在系统级别上，云边物协同技术还包含任务的迁移、资源隔离、任务调度等技术。在硬件级别上，该系统还包括专用芯片、硬件产品的设计与制造等，这方面的研究有很多相关的综述 [22]，本文不再赘述。

### 三、协同计算技术应用场景

针对目前边缘智能的典型场景，本文介绍三个协同技术应用的案例。三个案例分别是基于边云协同的个性化驾驶行为分析系统，基于边边协同的电动汽车电池故障检测系统和基于云边物协同的家电状态感知系统。

#### 基于边云协同的个性化驾驶行为分析系统

自动驾驶汽车是边缘计算的典型场景，其中 ADAS（Advanced Driving Assistant System，高级驾驶辅助系统）是实现自动驾驶的重要一环。针对驾驶行为建模是 ADAS 的重要组成部分，汽车公司需要根据其设计合适的辅助功能，保险公司也希望通过驾驶行为确定司机的保单等级。目前基于云端的模型只能衡量大众行为，而不同的驾驶员在不同外界环境下的驾驶行为是不同的，依赖于一个统一的模型并不合适。因此，Zhang 等人基于云边协同设计了一个个性化的驾驶行为分析系统 [23]。该模型根据汽车行驶时

[22]Sze V, Chen Y H, Yang T J, et al. Efficient processing of deep neural networks: A tutorial and survey[J]. Proceedings of the IEEE, 2017, 105(12): 2295–2329.

[23]Xingzhou Zhang, Mu Qiao, Liangkai Liu, Yunfei Xu and Weisong Shi, Collaborative Cloud-Edge Computation for Personalized Driving Behavior Modeling, in Proceedings of the fourth ACM/IEEE Symposium on Edge Computing (SEC), November 7–9, 2019, Arlington, VA, USA.

的经纬度、加速度、角速度等，对司机的驾驶行为建模，进而区分正常驾驶和异常驾驶行为（例如超速、醉驾等）。在该问题的研究中，如图 4 所示，云端首先根据大量数据训练通用模型，之后进行裁剪，再将模型迁移到边缘（车）上，最后根据边缘的个性化数据进行再次学习，从而使模型具有个性化和轻量级的特点。

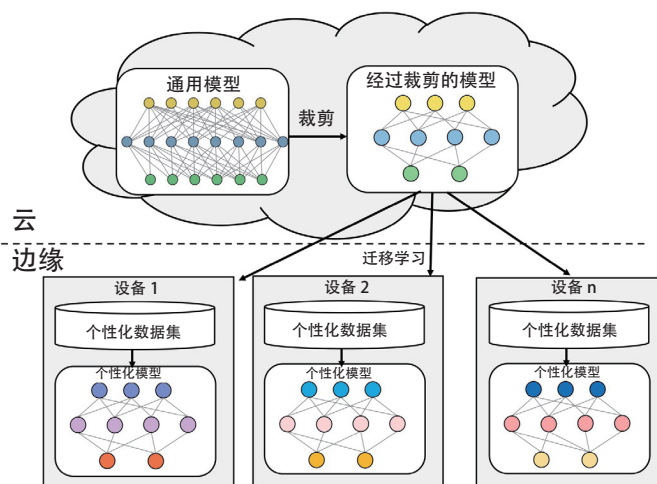


图 4：基于云边协同的个性化驾驶行为分析系统

### 基于边边协同的电动汽车电池故障检测系统

联网车辆能够更好地支持边缘计算。Lu 等人基于边缘与边缘之间相互协同的思想提出了 CLONE (Collaborative Learning on the Edges, 边缘协同学习) [24]，这是一个在边缘设备端实现协同合作的框架。如图 5 所示，CLONE 基于从大型电动汽车公司收集的真实数据集，选择电动汽车电池及相关配件的故障作为案例研究，以展示 CLONE 解决方案如何准确地预测故障，以协作的方式确保可持续、可靠的驾驶。CLONE 的提出建立在联邦学习算法和长短时记忆网络的基础之上，实现以车辆之间协同的方式同步训练个性化神经网络模型。在训练模型的过程中无需将任何车辆内部产生的真实数据集上传到云端，只需要将模型参数上传到参数边缘设备端，并在参数边缘设备端进行参数的加权整合，从而有效地防止私密信息泄露，保护用户隐私。同时，各个车辆采用异步通信的方式以减少延迟。实验结果表明，CLONE 可以在不降低模型精确

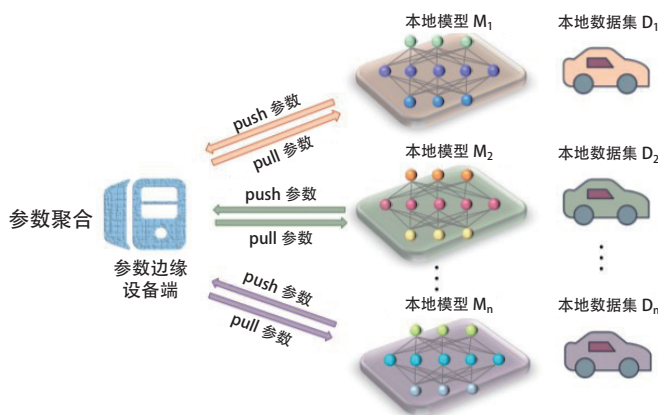


图 5：基于边边协同的电动汽车电池故障检测系统

[24]Lu S, Yao Y, Shi W. Collaborative Learning on the Edges: A Case Study on Connected Vehicles[C]//2nd {USENIX} Workshop on Hot Topics in Edge Computing (HotEdge 19). 2019.

度的情况下显著地减少训练模型的时间。

## 基于云边物协同的家电状态感知系统

智慧家庭是边缘智能的另一个典型场景。在智慧家庭中，实时感知家庭中家电的开关行为对用户行为分析和降低家庭整体功耗具有非常重要的意义。Zhang 等人设计了一个非侵入式的家电状态感知系统 [25] 以解决家电状态感知应用中功耗高和标记复杂的问题。该系统只使用一个传感器即可感知多个家电开关状态。具体而言，该系统利用了云边物协同的模式，如图 6 所示，物端（Main Current Sampling Sensor，主电流采样传感器）采集电流波形，该波形将会被传输到边缘端（Home Server，家庭服务器），边缘端运行机器学习算法。在云端存储一个大型的电器特征库，这些库与家庭电器的特征进行实时同步，并且对外提供服务。经过云边物协同后的计算系统，其功耗降至原来的 10% 以下，感知电器的数目由 5 个增加到 10 个以上，感知准确率达到 90% 以上。

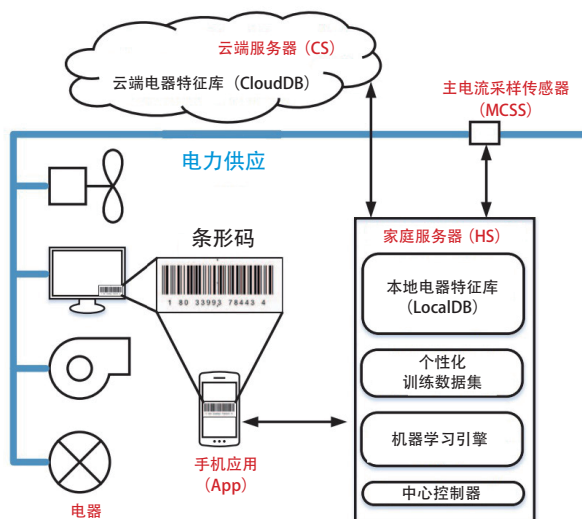


图 6: 基于云边物协同的家电状态感知系统

## 四、实现边缘智能中的协同需要面对的挑战

### 通信带宽的限制

在协同计算技术，不同设备间的通信问题是一个很大的挑战。在边缘智能场景下，不同设备之间存在地理位置的隔离，相互之间无法像云计算中心那样直接相连。不管是边缘与边缘之间共同训练 / 预测模型，还是边缘与物端之间协同完成一个任务，甚至是边缘与云之间进行协同分工，通信中在不同的计算层级之间进行协同时，带宽的限制都是一个挑战。5G 可能会成为解决通信问题的一个方案，但是也需要在算法层面进行优化，尽可能地减少通信的数据量，以保证服务质量。

### 缺乏统一的开发接口

在物联网的发展中，一直存在昆虫纲悖论。物联网设备的型号、类别、操作系统、

[25]Zhang X, Wang Y, Chao L, et al. IEHouse: A non-intrusive household appliance state recognition system[C]//2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2017: 1-8.



开发语言等众多，而边缘设备目前也存在这种趋势。尤其是边缘设备的异构性比较严重，在运行边缘智能应用时，既可以使用 CPU/GPU 等，也可以使用专用的加速器。缺乏统一的软硬件标准和统一的对外开发接口，会导致开发者编程困难，给协同计算的发展和应用的开发带来阻力。

### 缺乏跨平台边缘智能协同计算框架

目前，在边缘智能场景下，大多是基于云端已有的计算框架进行开发。然而，这些框架不完全适用于边缘场景，而且云边物设备的异构性强，这些框架不适用于多种设备的需求。框架的不统一降低了开发的效率，也无法充分发挥协同的优势。因此，设计通用的、可跨平台运行的计算框架有很大的必要性。

## 五、总结

人工智能应用在边缘设备上面临着诸多矛盾。利用边缘与云端、物端的协同发展，将有效地化解这些矛盾。本文对已有的协同模式进行分类并描述了每种协同所需要的关键技术和应用场景，最后提出为了实现协同技术所面临的几个挑战，包括通信带宽的限制、缺乏统一的开发接口和缺乏跨平台的边缘智能协同计算框架。相信随着协同计算技术和相应的软硬件技术的快速发展，边缘智能场景将越来越完善，催生出更多的人工智能场景落地。



查看内容精选