面向视频监控基于联邦学习的智能边缘计算技术

赵羽，杨洁，刘淼，孙金龙，桂冠
（南京邮电大学通信与信息工程学院，江苏 南京 210003）

摘 要：随着全球数据量的激增，集中式云计算无法提供低时延、高效率的视频监控服务。基于此，提出分布式边缘计算模型，在边缘端直接处理视频数据，减少网络的传输压力，缓解中央云服务器的计算负担，降低视频监控系统的处理时延。结合联邦学习算法，采用轻量级神经网络，分场景训练模型，并将其部署于计算能力受限的边缘设备上。实验结果表明，对比通用神经网络模型，所提方法检测准确度提高 18%，模型训练时间有效减少。

关键词：联邦学习；深度学习；边缘计算；轻量级神经网络；目标检测
中图分类号：TP391
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2020192

Federated learning based intelligent edge computing technique for video surveillance

ZHAO Yu, YANG Jie, LIU Miao, SUN Jinlong, GUI Guan
College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract: With the explosion of global data, centralized cloud computing cannot provide low-latency, high-efficiency video surveillance services. A distributed edge computing model was proposed, which directly processed video data at the edge node to reduce the transmission pressure of the network, eased the computational burden of the central cloud server, and reduced the processing delay of the video surveillance system. Combined with the federated learning algorithm, a lightweight neural network was used, which trained in different scenarios and deployed on edge devices with limited computing power. Experimental results show that, compared with the general neural network model, the detection accuracy of the proposed method is improved by 18%, and the model training time is reduced.

Key words: federated learning, deep learning, edge computing, lightweight neural network, object detection

1 引言

由于物联网技术[1]、云计算技术[2]以及第五代移动通信技术[3-4]的推动，全球数据量呈指数级增长。由中国信息通信研究院发表的《大数据白皮书》指出，全球数据量将于 2020 年年底达到 50 ZB（1 ZB=2^{40} GB）。面对如此庞大的数据量，传统的以云计算技术为核心的集中式数据处理方式逐渐显现出瓶颈。与此同时，随着边缘端硬件设备数据处理能力不断增强，以边缘计算技术[5-6]为核心的分布式数据处理方式应运而生。

视频监控作为物联网技术中的一环越来越受到人们的重视，它广泛应用于交通[7]、安防[8]等各个领域。早年间，人们通过建立专门的监控室来观测监控视频，这样做不但人力成本昂贵，而且不能及时捕捉关键信息。近些年，深度学习在目标检测领域快速发展，使智能视频监控系统在各个领域得以落地。但是面对海量的视频数据，传统的集中式云计算处理方式能力有限，主要体现在以下 3 个方面。

收稿日期：2020−04−22；**修回日期**：2020−07−07
基金项目：工信部重大专项基金资助项目（No.TC190A3WZ-2）；国家自然科学基金资助项目（No.61901228）
Foundation Items: The Major Project of the Ministry of Industry and Information Technology of China (No.TC190A3WZ-2), The National Natural Science Foundation of China (No.61901228)

1) 实时问题。随着边缘摄像头数量的不断增加，将在网络边缘产生大量的实时数据。把数据从边缘设备传输至云计算中心占用大量的网络带宽，造成网络时延。

2) 效率问题。在海量视频数据中，只有少部分数据具有信息价值，直接传送视频数据占用大量网络资源。基于深度学习的视频分析方法需要强大的计算能力支持，不经过处理的视频数据会给云计算中心带来计算负担。

3) 隐私问题。视频数据中携带有大量的个人隐私，将这些视频数据直接上传云计算中心增加了泄露用户隐私的风险。

近些年，为了解决深度学习过度依赖硬件资源的问题，研究者提出了多种轻量级神经网络模型。Zhang 等[9]提出分组卷积和通道随机组合操作对神经网络进行压缩，Howard 等[10]提出深度可分离卷积来减少标准卷积的运算量，Tan 等[11]提出自动神经网络结构体系搜索方法来构建轻量级神经网络。这些方法使神经网络可以在移动终端或者嵌入式设备中运行。吕华章等[12]介绍了近些年边缘计算标准化进展情况，张佳乐等[13]探究了边缘计算数据安全与隐私保护现状。大量参考文献表明，边缘计算可以有效突破云计算当前瓶颈，是大势所趋。

基于上述考虑，本文提出一种基于联邦学习（FL, federated learning）的分布式边缘计算模型，并将其应用于视频监控系统，该方法仅传输神经网络模型权重，免于视频数据的传输。经实验发现，本文提出的分布式边缘计算模型可以针对不同场景实现单独训练，以提高目标检测的准确性。联邦学习的引入不仅可以保证边缘端用户隐私不会泄露，还可以减少神经网络模型的训练时间。

2 系统模型

2.1 基于联邦学习的边缘计算视频监控模型

图 1 是传统的集中式云计算模型。在当前行情下，不同企业拥有不同数据格式的视频数据库，集中式云计算模型需要将这些数据统一上传至云端服务器，经过数据清洗和数据格式统一后训练出一个通用的神经网络检测模型。在完成神经网络模型的建立后，各企业需将待检测的视频上传至云端服务器统一检测，服务器再将检测结果逐一下发。这样会占用大量的网络资源，且难以满足一些对实时性要求较高的应用场景。各个企业之间存在数据壁垒，将数据上传至云端服务器存在数据泄露的风险。除此之外，由于各企业数据量不同，通用的神经网络模型对于数据量较少的企业不友好，检测效果较差。

针对以上问题，微众银行[14]提出如图 2 所示的联邦学习模型。在联邦学习模型中，各企业只需将神经网络模型下载到本地，利用本地的数据进行训练，最后将训练完成的权重参数上传。云端服务器根据各企业上传的权重参数联合优化神经网络模型。视频的检测与分析都在数据库本地进行，既保证了数据的安全性，也保证了数据处理的实时性。

本文结合云计算模型与联邦学习模型，提出基于联邦学习的边缘计算视频监控模型，如图 3 所示。该模型由 4 个部分组成，分别是云端服务器、企业服务器、边缘开发板和摄像头。云端服务器负责存储公共数据集和训练通用神经网络模型，并将网络模型分发给企业服务器；企业服务器针对不同场景根据本地数据库训练与更新神经网络模型，并将算法部署于边缘开发板 NVIDIA TX2 上；边缘开发板

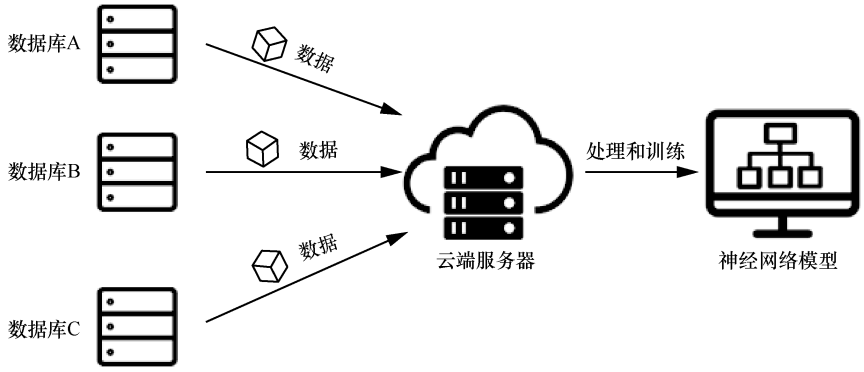


图 1 传统的集中式云计算模型

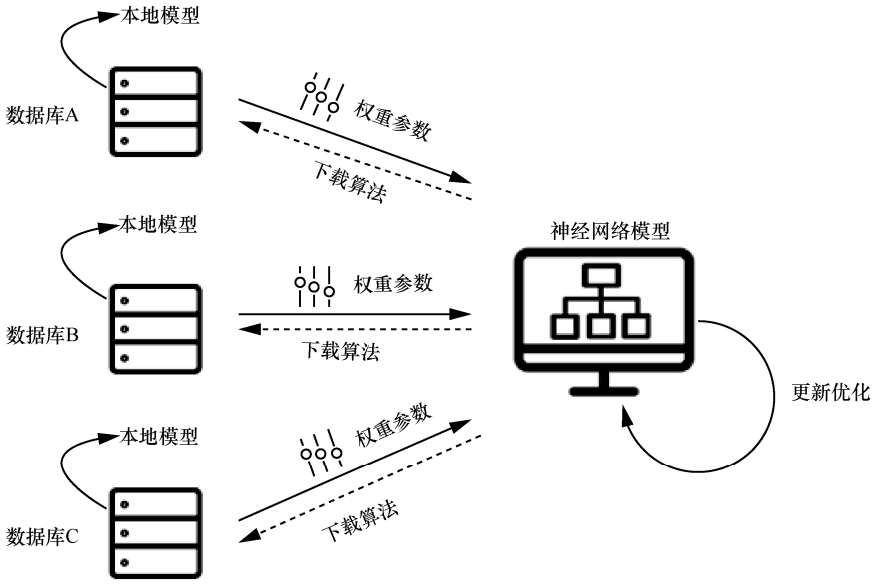


图 2　联邦学习模型

NVIDIA TX2 负责实时分析视频数据，并将处理后的视频数据上传至企业服务器统一管理；摄像头负责实时采集视频。企业服务器完成神经网络模型训练后，将网络权重上传至云端服务器，云端服务器联合处理各企业的网络权重，用于更新通用神经网络模型。

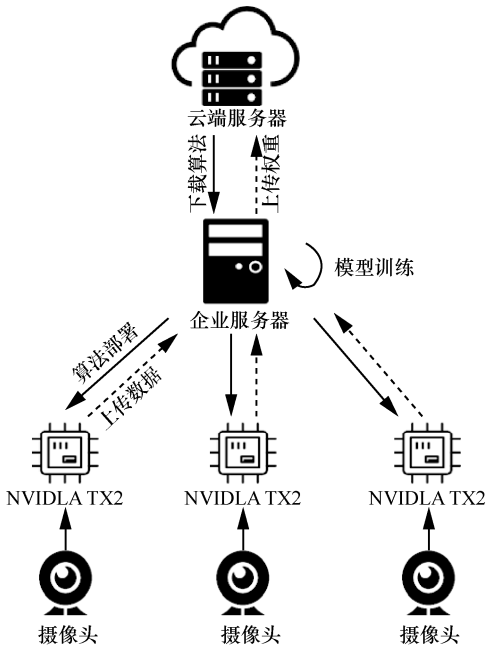


图 3　基于联邦学习的边缘计算视频监控模型

本文以智能安全帽检测系统为例来验证基于联邦学习的边缘计算视频监控模型。为了提高神经网络在边缘开发板上的检测速度，本文采用改进 YOLOv3[15]神经网络检测模型。

2.2　改进 YOLOv3 神经网络模型

在基于深度学习的目标检测领域，根据检测方式的不同，检测方法大体可以分为两类：基于候选区域的两步目标检测算法和基于端到端的一步目标检测算法。其中，基于端到端的一步目标检测算法以其检测速度快著称。本文选取基于端到端的目标检测算法 YOLOv3，并在其基础上进行改进，使其能够在计算能力受限的边缘设备 NVIDAI TX2 上快速运行。

YOLOv3 的检测原理为将输入的图像人为地划分为 T^2 个方格，每小方格产生 A 个边界框，若待检测物体的中心位置位于某个小方格中，则由该小方格负责预测该物体。改进 YOLOv3 神经网络结构如图 4 所示。

改进 YOLOv3 神经网络结构主要分为 4 个部分：输入层、主网络框架、多尺度检测模块和非极大值抑制模块。主网络框架由卷积层组合而成，用于提取图像特征。与 YOLOv3 不同的是，本文采用文献[10]提出的深度可分离卷积层代替标准卷积层，通过将标准卷积分解为深度卷积和逐点卷积，从而降低了神经网络的浮点数运算量，其结构对比如图 5 所示。多尺度检测模块从 3 个不同尺度的特征图上检测物体，其中大尺度的特征图含有更多特征信息，负责检测小物体；小尺度的特征图含有更多全局信息，负责检测大物体。非极大值抑制模块用于获取局部最大值，从而抑制神经网络产生多余的候选框。

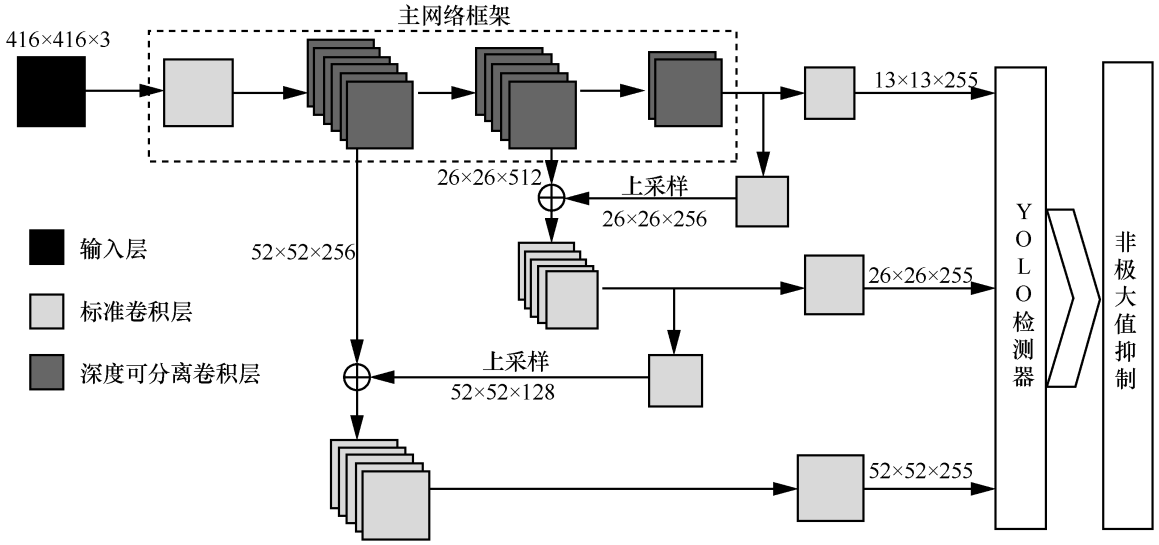
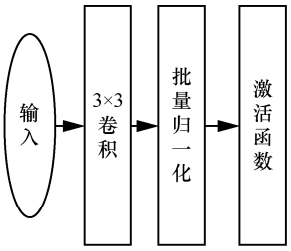
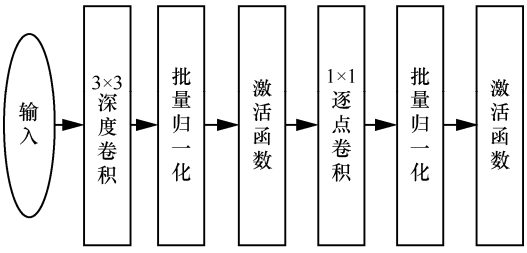


图 4 改进 YOLOV3 神经网络结构



(a) 标准卷积层结构



(b) 深度可分离卷积层结构

图 5 标准卷积与深度可分离卷积结构对比

改进 YOLOv3 的损失函数如式(1)所示，主要由中心坐标损失 $\text{loss}_{\text{center}}$、边界框大小损失 $\text{loss}_{\text{scale}}$、置信度损失 $\text{loss}_{\text{conf}}$ 以及分类损失 $\text{loss}_{\text{class}}$ 这 4 个部分组成。$\text{loss}_{\text{center}}$ 和 $\text{loss}_{\text{scale}}$ 采用和方差的计算方式，如式(2)和式(3)所示；$\text{loss}_{\text{conf}}$ 和 $\text{loss}_{\text{class}}$ 采用交叉熵的计算方式，如式(4)和式(5)所示。

$$\text{loss}=\text{loss}_{\text{center}}+\text{loss}_{\text{scale}}+\text{loss}_{\text{conf}}+\text{loss}_{\text{class}} \tag{1}$$

$$\text{loss}_{\text{center}}=\sum_{i=0}^{T^2}\sum_{j=0}^{A}I_{ij}\left[\left(\hat{x}_i^j-x_i^j\right)^2+\left(\hat{y}_i^j-y_i^j\right)^2\right] \tag{2}$$

$$\text{loss}_{\text{scale}}=\sum_{i=0}^{T^2}\sum_{j=0}^{A}I_{ij}\left[\left(\sqrt{\hat{w}_i^j}-\sqrt{w_i^j}\right)^2+\left(\sqrt{\hat{h}_i^j}-\sqrt{h_i^j}\right)^2\right] \tag{3}$$

$$\text{loss}_{\text{conf}}=-\sum_{i=0}^{T^2}\sum_{j=0}^{A}I_{ij}\left[\hat{C}_i^j\ln(C_i^j)+(1-\hat{C}_i^j)\ln(1-C_i^j)\right] \tag{4}$$

$$\text{loss}_{\text{class}}=-\sum_{i=0}^{T^2}\sum_{j=0}^{A}I_{ij}\left[\hat{P}_i^j\ln(P_i^j)+(1-\hat{P}_i^j)\ln(1-P_i^j)\right] \tag{5}$$

其中，$I_{ij}\in\{0,1\}$ 为判断第 i 个方格的第 j 个边界框是否包含待检测物体；(x_i^j,y_i^j) 为真实物体的中心点坐标；(w_i^j,h_i^j) 为真实物体边界框的宽和高；C_i^j 为真实物体的置信度；P_i^j 为真实物体的类别置信度；$\hat{x}_i^j$、$\hat{y}_i^j$、$\hat{w}_i^j$、$\hat{h}_i^j$、$\hat{C}_i^j$、$\hat{P}_i^j$ 分别为 x_i^j、y_i^j、w_i^j、h_i^j、C_i^j、P_i^j 的预估值。表 1 为 YOLOv3 与改进 YOLOv3 在网络大小、十亿浮点数计算量（BFLOP, billion floating point operations）、公共数据集 PASCAL VOC 中的平均精度均值（mAP, mean average precision）以及推理速度的比较。

表 1 YOLOv3 与改进 YOLOv3 对比

比较项	YOLOv3	改进 YOLOv3
网络大小/MB	171	37
BFLOP	65.73	15.65
mAP	81.5%	76.82%
推理速度/ms	56	19

2.3 通用模型更新方法

假设 W^0 为由公共数据集训练的初始神经网络权重，$\{W_i\}_{i=1}^n$ 为不同场景根据本地数据集训练的神经网络权重，$\{\lambda_i\}_{i=1}^n$ 为不同场景数据集数据量的衡量尺度，$\overline{W}$ 为通用神经网络权重，其更新方法如式(6)所示。

$$\overline{W}=\frac{\sum_{i=1}^{n}\lambda_i W_i}{\sum_{i=1}^{n}\lambda_i} \tag{6}$$

考虑到边缘端数据集数据量较小的问题，为了防止神经网络模型过拟合，$\{W_i\}_{i=1}^{n}$ 采用迁移学习[16]的方式训练。企业服务器在接收到来自云端服务器的通用神经网络模型后，固定主网络框架的权重，根据本地数据集微调模型即可，这样可以极大地缩短神经网络模型的训练时间。

3　实验与结果分析

在本实验中，企业服务器采用的图形处理单元（GPU, graphics processing unit）为 NVIDIA GTX 1080Ti，边缘开发板为 NVIDIA TX2，操作系统为 Ubuntu 16.04，编程语言为 Python，深度学习框架为 TensorFlow。表 2 为企业服务器与 NVIDIA TX2 的性能、功耗、成本对比。

表 2　企业服务器与 NVIDIA TX2 的性能、功耗、成本对比

比较项	企业服务器	NVIDIA TX2
CUDA 核心/个	3 584	256
运行内存/GB	128	8
功耗/W	600	7.5
硬件成本/元	30 000	3 150

3.1　数据集的建立

在云端服务器侧，本实验通过网络爬虫以及人工标注的方式，结合通用安全帽佩戴检测数据集（SHWD, safety helmet wearing detect dataset）制作了包含 15 093 幅图像的数据集，用于训练通用神经网络。在企业服务侧，根据摄像头角度以及距离远近，本实验制作了 3 组不同场景的数据集，每组数据集包含 2 000 多幅图像，通过迁移学习的方式验证不同场景下基于联邦学习的边缘计算视频监控系统的检测准确度。为了丰富数据集，提高检测效果，在训练神经网络模型之前，本实验采用了数据增强技术。除了常见的翻转、按比例缩放、随机裁剪、移位以及添加高斯噪声等数据增强方法，本实验还采用了 Mixup[17]数据增强方法。

Mixup 原理如图 6 所示。将 2 幅图像缩放至同一大小后逐像素相加并按比例融合，2 幅图像的标签组合成新的标签。在计算损失函数时，需按照融合时的比例分别计算损失函数再相加。此种方式通过人为地引入遮挡来提高目标检测的准确性，尤其适用于安全帽检测这种含有目标遮挡的场景。

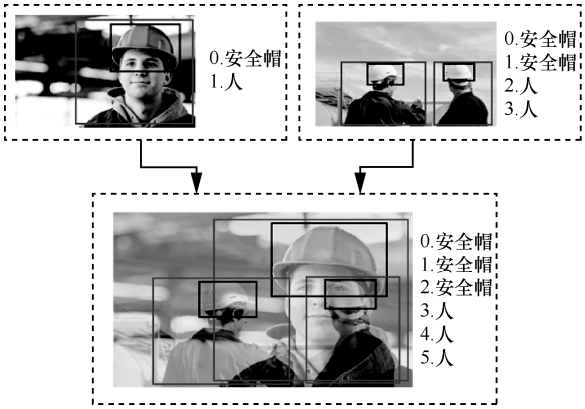


图 6　Mixup 原理

3.2　评价指标

在多分类目标检测任务中，神经网络检测模型的好坏并不能单纯由精准率来衡量，需要综合评定模型的分类和定位性能。在目标检测领域，定位性能依靠交并比（IoU, intersection over union）来衡量，即预测边界框与真实边界框的交集和并集的比值。在本实验中，当 IoU≥0.5 时判定为正确检测。对于某一特定类别，根据 IoU 以及置信度阈值可以计算出精准率与召回率，并进一步计算平均精度（AP, average precision），最后计算各类别 AP 的平均值即为 mAP。

3.3　结果分析

图 7 为 3 个不同场景数据集训练改进 YOLOv3 与 MobileNet-SSD[10]时的损失值与 mAP 对比，其中 mAP 每遍历一次数据集计算一次，损失值取值为遍历数据集时每次神经网络权重更新时的平均值。由于改进 YOLOv3 与 MobileNet-SSD 两者的损失函数定义不同，无法从损失值直接观察出模型的检测准确度，但可以得出，两者均需要遍历 160 次左右数据集来完成模型训练，它们的 mAP 对比如表 3 所示。

实验发现，若采用集中式云计算模型，仅训练一个云端通用神经网络，测得 3 个数据集的 mAP 分别为 71.98%、68.32%和 73.05%。除此之外，在实际应用中，企业服务器处理每帧图像的时间为 19 ms，若采用集中式云计算模型，系统无法同时处理多路摄像头的视频数据，大量的待检测视频数据会导致系统内存崩溃。通过迁移学习的方式分场景训练，改进 YOLOv3 神经网络较云端通用神经网络 mAP

平均提升 18%，并且由于本地数据集数据量较少，神经网络的训练时间大大减少。企业服务器在完成神经网络模型的训练后将算法部署于边缘开发板 NVIDIA TX2 上，从而实现视频数据的本地化实时处理。NVIDA TX2 支持统一计算设备架构（CUDA, compute unified device architecture）与 TensorFlow 深度学习框架，不需要模型转换即可直接运行神经网络模型，检测速度可达 18 frame/s，其硬件成本仅为企业服务器的十分之一。最后，企业服务器上传神经网络权重至云端服务器用于更新通用神经网络。图 8 为安全帽检测实际效果。

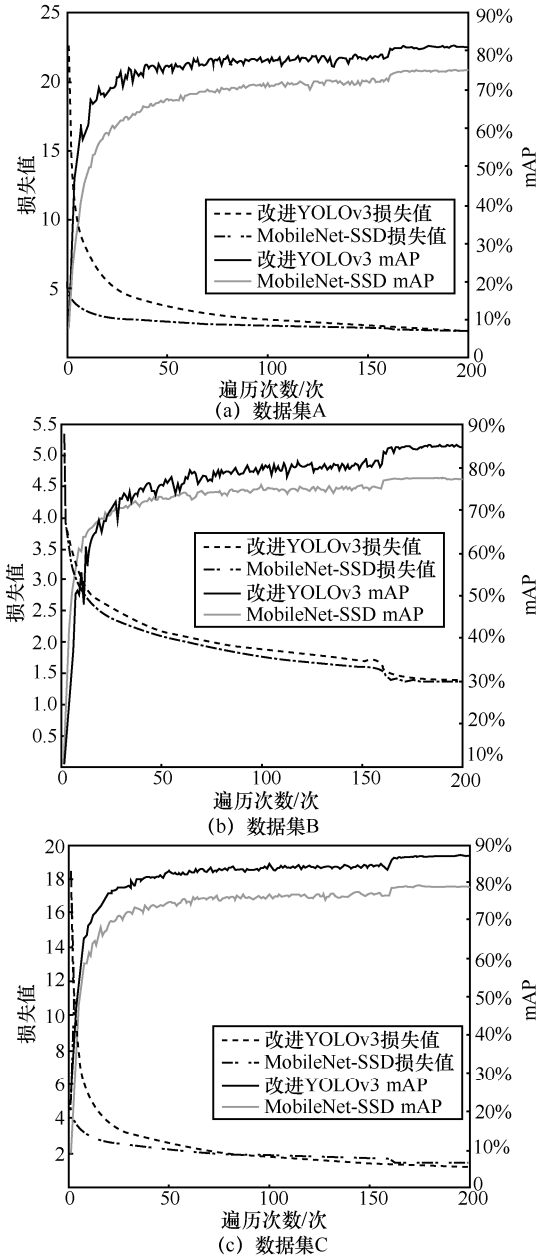


图 7　改进 YOLOv3 与 MobileNet-SSD 的损失值与 mAP

表 3　改进 YOLOv3 与 MobileNet-SSD 的 mAP 对比

数据集	MobileNet-SSD	改进 YOLOv3
数据集 A	75.38%	81.31%
数据集 B	77.44%	85.19%
数据集 C	79.19%	87.35%

图 8　安全帽检测实际效果

4　结束语

现有的集中式云计算处理视频监控的方式网络数据传输量大，系统时延高且可能泄露用户隐私。针对这些问题，本文提出基于联邦学习的边缘计算视频监控系统，采用轻量级神经网络模型，将视频的分析处理分布于边缘开发板中，神经网络模型的训练分布于企业服务器中，有效地缓解云端服务器的计算与存储负担，免去视频数据的传输，并且在打破数据孤岛的同时保护用户隐私。除此之外，本文所提的基于联邦学习的边缘计算视频监控系统相较于传统系统节约了硬件成本以及运行功耗，更有利于实际项目的部署。未来工作将继续研究联邦学习算法，优化通用神经网络权重的更新方式，进一步提升检测效果，引入神经网络压缩与加速算法，使系统达到实时要求。

参考文献：

[1] Al-FUQAHA A, GUIZANI M, MOHAMMADI M. Internet of things: a survey on enabling technologies, protocols, and applications[J] IEEE Communications Surveys & Tutorials, 2015, 17(4): 2347-2376.

[2] SADOOGHI I, MARTIN J H, LI T L. Understanding the performance and potential of cloud computing for scientific applications[J]. IEEE Transactions on Cloud Computing, 2017, 5(2): 358-371.

[3] 张平, 陶运铮, 张治. 5G 若干关键技术评述[J]. 通信学报, 2016,

37(7): 15-29.
ZHANG P, TAO Y Z, ZHANG Z. Survey of several key technologies for 5G[J]. Journal on Communications, 2016, 37(7): 15-29.

[4] 桂冠, 王禹, 黄浩. 基于深度学习的物理层无线通信技术: 机遇与挑战[J]. 通信学报, 2019, 40(2): 19-23.
GUI G, WANG Y, HUANG H. Deep learning based physical layer wireless communication techniques: opportunities and challenges[J]. Journal on Communications, 2019, 40(2): 19-23.

[5] 施巍松, 孙辉, 曹杰. 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, 54(5): 907-924.
SHI W S, SUN H, CAO J. Edge computing: an emerging computing model for the internet of everything era[J]. Journal of Computer Research and Development, 2017, 54(5): 907-924.

[6] SHI W S, CAO J, ZHANG Q. Edge computing: vision and challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646.

[7] MHALLA A, CHATEAU T, GAZZAH S. An embedded computer-vision system for multi-object detection in traffic surveillance[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(11): 4006-4018.

[8] HU L, NI Q. IoT-driven automated object detection algorithm for urban surveillance systems in smart cities[J]. IEEE Internet of Things Journal, 2018, 5(2): 747-754.

[9] ZHANG X Y, ZHOU X Y, LIN M X. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2018: 6848-6856.

[10] HOWARD A, ZHU M L, CHEN B. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv Preprint, arXiv:1704.04861, 2017.

[11] TAN M X, CHEN B, PANG R M. MnasNet: platform-aware neural architecture search for mobile[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2019: 2815-2823.

[12] 吕华章, 陈丹, 范斌. 边缘计算标准化进展与案例分析[J]. 计算机研究与发展, 2018, 55(3): 487-511.
LYU H Z, CHEN D, FAN B. Standardization progress and case analysis of edge computing[J]. Journal of Computer Research and Development, 2018, 55(3): 487-511.

[13] 张佳乐, 赵彦超, 陈兵. 边缘计算数据安全与隐私保护研究综述[J]. 通信学报, 2018, 39(3): 1-21.
ZHANG J L, ZHAO Y C, CHEN B. Survey on data security and privacy-preserving for the research of edge computing[J]. Journal on Communications, 2018, 39(3): 1-21.

[14] YANG Q, LIU Y, CHEN T. Federated machine learning: concept and applications[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-9.

[15] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv Preprint, arXiv: 1804.02767, 2018.

[16] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[17] ZHANG Z, HE T, ZHANG H. Bag of freebies for training object detection neural networks[J]. arXiv Preprint, arXiv: 1902.04103, 2019.

[作者简介]

赵羽（1996− ），男，江苏泰州人，南京邮电大学博士生，主要研究方向为基于深度学习的智能视频分析。

杨洁（1980− ），女，江苏南京人，博士，南京邮电大学讲师，主要研究方向为基于深度学习的智能视频分析。

刘淼（1988− ），男，江苏淮安人，博士，南京邮电大学讲师，主要研究方向为基于深度学习的智能通信。

孙金龙（1988− ），男，河南洛阳人，博士，南京邮电大学讲师，主要研究方向为基于深度学习的智能通信。

桂冠（1982− ），男，安徽枞阳人，博士，南京邮电大学教授，主要研究方向为基于深度学习的物理层无线通信技术。