DOI: 10.1609/aaai.12002

SPECIAL TOPIC ARTICLE

Federated learning-powered visual object detection for safety monitoring

Yang Liu[1] | Anbu Huang[1] | Yun Luo[2,3] | He Huang[3] | Youzhi Liu[1] | Yuanyuan Chen[4] | Lican Feng[3] | Tianjian Chen[1] | Han Yu[4,5] | Qiang Yang[1,2]

[1] Department of AI, WeBank, Shenzhen, China

[2] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong

[3] Extreme Vision Ltd, Shenzhen, China

[4] School of Computer Science and Engineering, Nanyang Technological University, Singapore

[5] Joint NTU-WeBank Research Centre on FinTech, Singapore

Correspondence
Yang Liu, Department of AI, WeBank, Shenzhen, China.
Email: yangliu@webank.com

Yun Luo, Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong.
Email: lauren.luo@extremevision.mo

Han Yu, Extreme Vision Ltd, Shenzhen, China.
Email: han.yu@ntu.edu.sg

Abstract

Visual object detection is an important artificial intelligence (AI) technique for safety monitoring applications. Current approaches for building visual object detection models require large and well-labeled dataset stored by a centralized entity. This not only poses privacy concerns under the General Data Protection Regulation (GDPR), but also incurs large transmission and storage overhead. Federated learning (FL) is a promising machine learning paradigm to address these challenges. In this paper, we report on *FedVision*—a machine learning engineering platform to support the development of federated learning powered computer vision applications—to bridge this important gap. The platform has been deployed through collaboration between WeBank and Extreme Vision to help customers develop computer vision-based safety monitoring solutions in smart city applications. Through actual usage, it has demonstrated significant efficiency improvement and cost reduction while fulfilling privacy-preservation requirements (e.g., reducing communication overhead for one company by 50 fold and saving close to 40,000RMB of network cost per annum). To the best of our knowledge, this is the first practical application of FL in computer vision-based tasks.

INTRODUCTION

Visual object detection is one of the most important artificial intelligence (AI) techniques with wide applications in safety monitoring. As deep learning techniques advances, visual object detection has also witnessed significant development in recent years (Redmon and Farhadi 2018; Ren et al. 2017; Zhao et al. 2018). The current visual object detection model training approach requires centralized storage of training data (Figure 1). Under such an approach, each user annotates visual data from locally owned cameras and uploads these labeled training data to a central server (e.g., a cloud server). Data storage and model training both take place in the server.

Under such a machine learning paradigm, the users have no control over how the data would be used once they are transmitted to the central database, which makes them vulnerable to privacy breach. Besides, it is also difficult to share data across organizations due to liability concerns, which are made even more pronounced by data privacy protection regulations such as the General Data Protection Regulation (GDPR) (Voigt and Bussche 2017). The typically amount of data required to train a useful visual object detector also means that significant communication cost is incurred when transmitting training data from their sources to the server.

These challenges have motivated the AI research community to seek new paradigms of training machine learning models. Federated Learning (FL) (Kairouz et al. 2019; Yang et al. 2019) is one promising paradigm. Under FL, machine learning models are trained from distributed datasets without requiring data to leave their sources.

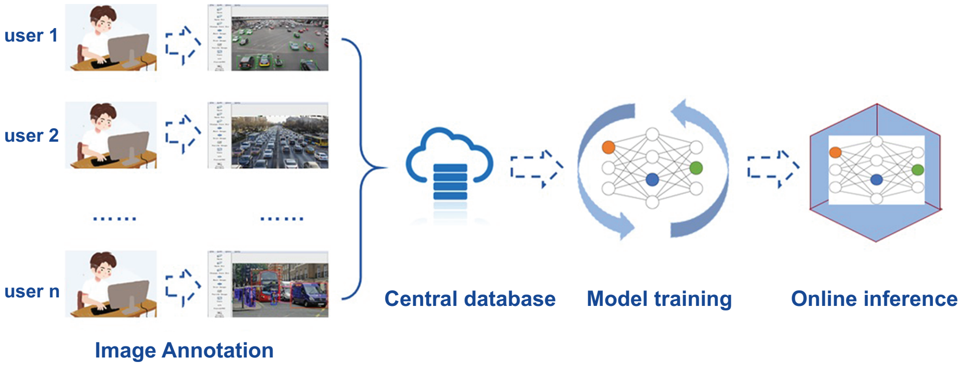


FIGURE 1 A typical workflow for centralized training of a visual object detector

Such an approach fundamentally limits data privacy leakage and communication overhead, balancing performance and efficiency issues while preventing sensitive data from being disclosed. FL models are trained through model aggregation rather than data aggregation. Under such a paradigm, we can train a visual object detection model locally at a data owner's site, and upload the model parameters to a central server for aggregation.

However, realizing this vision requires an easy to use tool to enable developers who are not experts in federated learning to conveniently leverage this technology to develop computer vision applications. Such a tool is not yet available. In order to bridge this gap, we propose *FedVision* (Liu et al. 2020). It is a machine learning engineering platform which supports easy development of FL-powered computer vision applications. The current version of FedVision supports a proprietary federated visual object detection algorithm framework based on YOLOv3 (Redmon and Farhadi 2018) – FedYOLOv3. It allows end-to-end collaborative training of FedYOLOv3 with locally stored datasets from multiple clients in a user-friendly manner.

The platform was deployed through collaboration between *WeBank*[1] and *Extreme Vision*[2] in May 2019. So far, it has been adopted by three large-scale corporate customers to develop visual object detection applications for safety monitoring. Through actual usage, the platform has helped the customers significantly improve their operational efficiency and reduce their costs, while eliminating the need to transmit sensitive data around (e.g., reducing communication overhead for one company by 50 fold and saving close to 40,000 RMB of network cost per annum). To the best of our knowledge, this is the first industry application of federated learning in computer vision-based tasks.

SYSTEM DESIGN

Under FedVision, training a visual object detection model consists of three main steps: (1) crowdsourcing image annotations, (2) federated model training, and (3) federated model update. In this section, we describe the design of these steps in detail.

Crowdsourcing image annotations

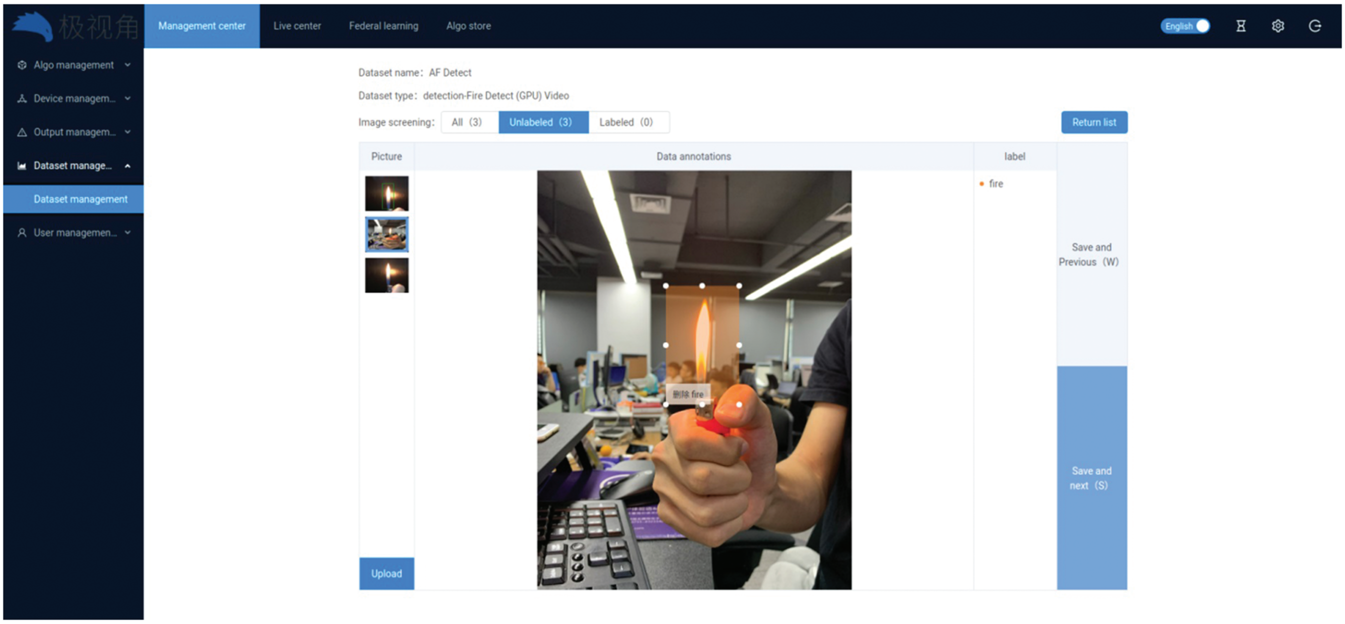
FedVision adopts the Darknet model format[3] for annotation. Under this format, each bounding box is specified as {label x y w h}, where "label'" denotes the category of objects, (x, y) represents the center of the bounding box, and (w, h) represents the width and height of the bounding box. This module provides data owners with a tool to easily label their locally stored image data for FL model training (Figure 2). Through the image annotation tool, a user can easily specify bounding boxes and the corresponding labels.

Anyone capable of visually identify where the objects of interest (e.g., flames) are in a given image can use the mouse to draw the bounding box, and assign it to a category/label. Users are not required to be familiar with federated learning. With this tool, the task of labeling training data can be distributed among data owners in a way similar to crowdsourcing (Doan, Ramakrishnan, and Halevy 2011), thereby, making it flexible to involve additional manpower in order to spread out the burden of image annotation. It also supports online learning with new image data arrive sequentially over time from the cameras.

Federated model training

The federated model training framework under FedVision is a variant of horizontal federated learning (HFL) (Yang et al. 2019). HFL, also known as sample-based federated learning, is designed for scenarios in which datasets have significant overlap in the feature space, but little overlap in the sample space (Figure 3).

HFL is suitable for the application scenario of FedVision since it helps multiple data owners with data from the



Dataset name: AF Detect

Dataset type: detection-Fire Detect (GPU) Video

Image screening: All (3) Unlabeled (3) Labeled (0)

FIGURE 2 The image annotation module of FedVision

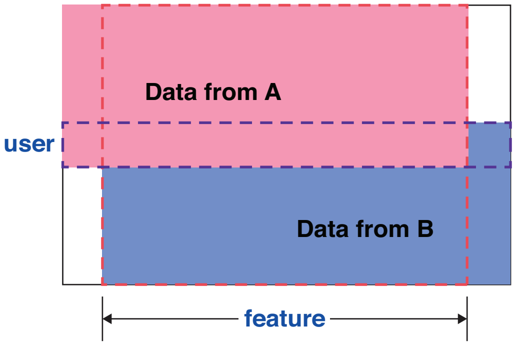


FIGURE 3 The concept of horizontal federated learning (HFL)

same feature space (i.e. labeled image data) to jointly train federated object detection models. The term “horizontal” comes from “horizontal partition”, which is widely used to describe the traditional tabular view of a database (i.e. rows of a table are horizontally partitioned into different groups and each row contains the complete set of data features).

Under HFL, data collected and stored by each party are no longer required to be uploaded to a common server to facilitate model training. Instead, the model framework is sent from the federated learning server to each party, which then uses the locally stored data to train this model. After training converges, the encrypted model parameters from each party are sent back to the server. They are then aggregated into a global model. This global model will eventually be distributed to the FL participants to be used for inference tasks.

From a system architecture perspective, the federated model training module consists of the following six components as shown in Figure 4:

Configuration: which allows users to configure model training by specifying the number of iterations, the number of reconnections, the server URL for uploading model parameters, and other key parameters.

Task Scheduler: which performs global dispatch scheduling to coordinate communications between the federated learning server and the clients in order to balance the utilization of local computational resources during the federated model training process. The load-balancing approach is based on (Yu et al. 2017) which jointly considers clients’ local model quality and the current load on their local computational resources in an effort to maximize the quality of the resulting federated model.

Task Manager: which coordinates the concurrent federated model training processes when multiple model algorithms are being trained concurrently by the FL clients.

Explorer: which monitors the resource utilization situation on the client side (e.g., CPU usage, memory usage, network load, etc.), so as to inform the Task Scheduler on its load-balancing decisions.

FL_SERVER: which is responsible for model parameter uploading, model aggregation and model dispatch, which are essential steps involved in federated learning (Bonawitz et al. 2019).

FL_CLIENT: which hosts the Task Manager and Explorer components and performs local model training, which is also an essential step involved in federated learning (Bonawitz et al. 2019).

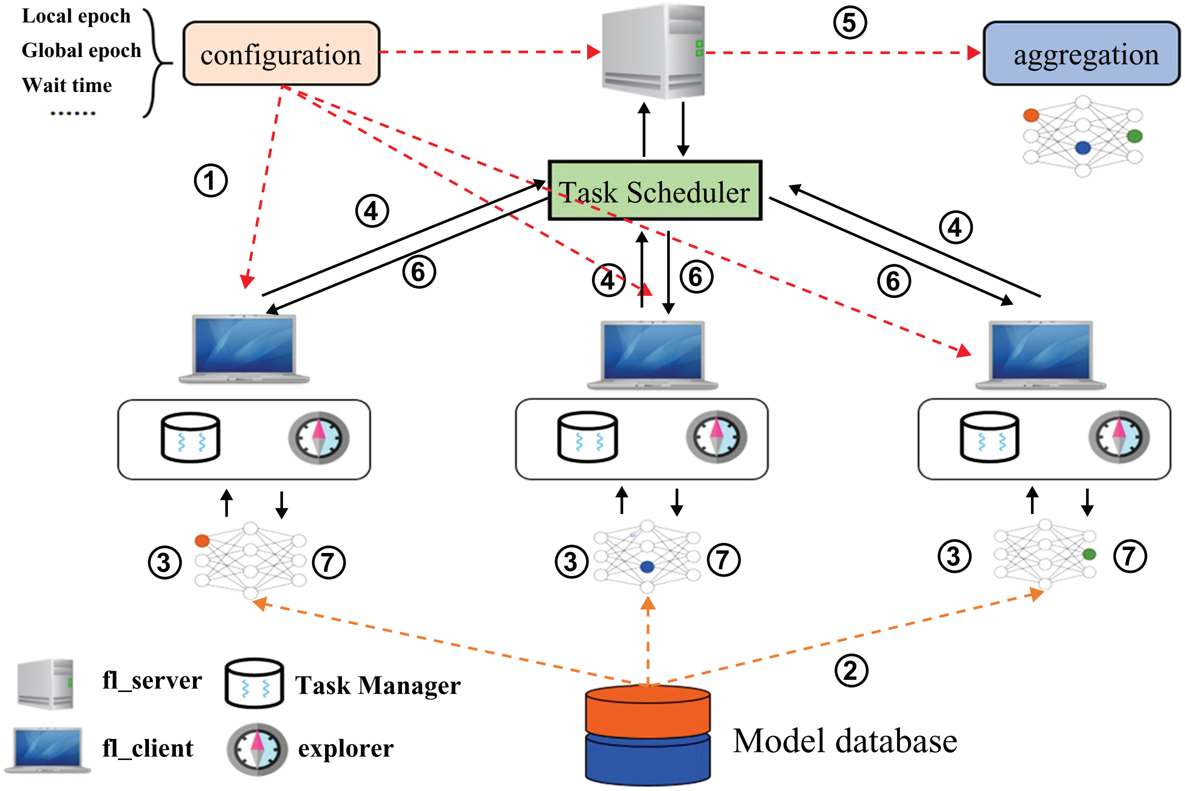


FIGURE 4 The system architecture of the federated model training module

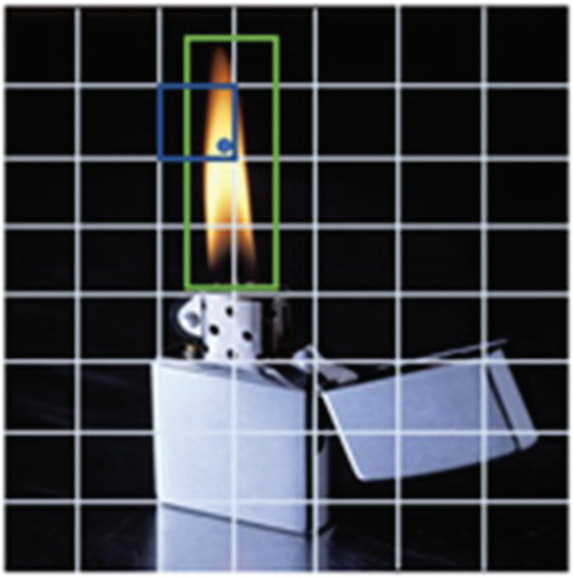


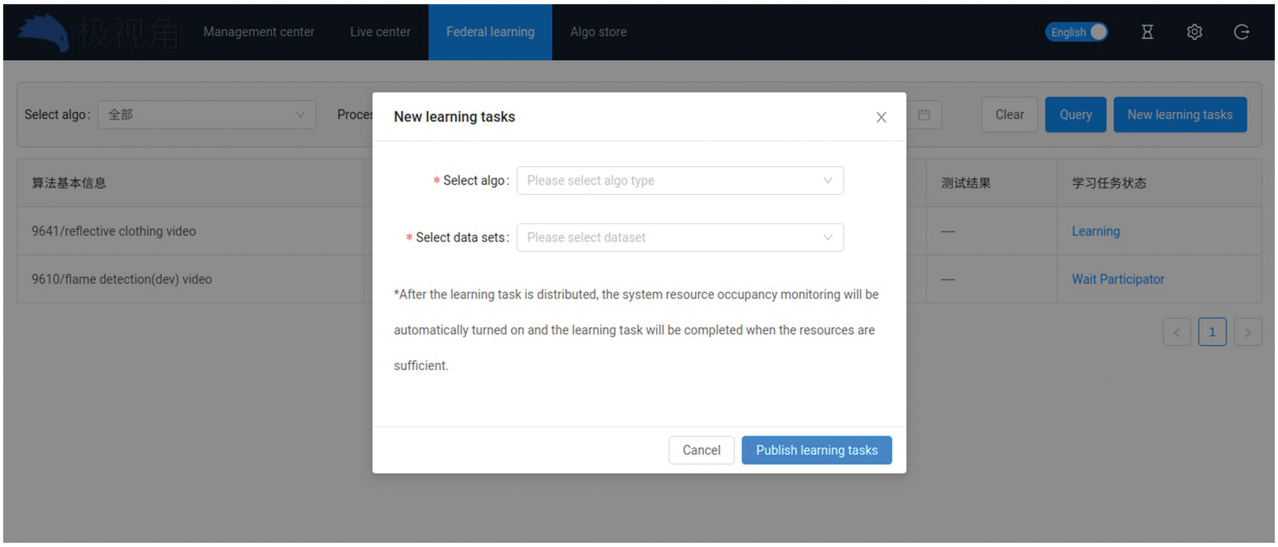
FIGURE 5 Flame detection with YOLOv3

As the application scenarios for FedVision places more emphasis on efficiency over performance, we adopt a one-stage approach YOLOv3 (Redmon and Farhadi 2018) as the basic object detection model. The approach of YOLOv3 can be summarized as follows. Given an image (e.g., Figure 5), it is first divided into an S x S grid with each grid square used for detecting the target object with its center located in that grid square (i.e. the blue square grid in Figure 5 is used to detect flames).

For each grid square, YOLOv3 predicts the positions of the bounding boxes, estimates the confidence score for the predicted bounding boxes, and computes the class conditional probability. The loss function of YOLOv3 consists of three parts: (1) class prediction loss, (2) bounding box coordinate prediction loss, and (3) confidence score prediction loss.

We implement a federated learning version of it—*Federated YOLOv3 (FedYOLOv3)*—in our platform. With one round of end-to-end training, FedYOLOv3 can identify the position of the bounding box as well as the class for the target object in an image.

After the users have labeled their local training datasets with the FedVision image annotation tool, they can join the FedYOLOv3 model training process. Once the local model converges, a user can initiate the transfer of the current local model parameters (in the form of a weight matrix) to the FL_SERVER in a secure encrypted manner through his FL_CLIENT. The HFL module in FedVision operates in rounds. After each round of learning elapses, FL_SERVER performs federated averaging (McMahan et al. 2016) to compute an updated global weight matrix for the model. The FL_SERVER then sends the updated



New learning tasks

* Select algo : Please select algo type

* Select data sets : Please select dataset

*After the learning task is distributed, the system resource occupancy monitoring will be automatically turned on and the learning task will be completed when the resources are sufficient.

算法基本信息	测试结果	学习任务状态
9641/reflective clothing video	---	Learning
9610/flame detection(dev) video	---	Wait Participator

FIGURE 6 Monitoring multiple rounds of federated model training on FedVision

weight matrix to the participating FL_CLIENTs so that they can enjoy the benefits of an updated object detection model trained with everyone's dataset in essence. Multiple rounds of FL model training can be monitored in the FedVision tool user interface as shown in Figure 6.

FEDERATED MODEL UPDATE

After local model training, the model parameters from each user are transmitted to the FL_SERVER by the respective FL_CLIENT. The updated model parameters need to be stored for tracing purposes. Over time, the storage size required for these parameter files increases with the rounds of FL model training. FedVision adopts Cloud Object Storage (COS) to address the problem of dynamic changes in the required storage space.

Transmitting model parameters from FL-CLIENTs to the FL_SERVER can be time consuming due to network bandwidth constraints. During the federated model training, different model parameters might have different contributions towards final model performance. Thus, neural network compression can be performed to reduce the sizes of the transmitted model parameters by pruning less useful weight values while preserving model performance (Bengio and LeCun 2016; Cheng et al. 2017). In FedVision, we apply network pruning to compress the federated model parameters and speed up transmission.

Let $M^{i,k}$ be the model parameter matrix from the i-th user after completing the k-th iteration of federated model training. Let $M_j^{i,k}$ be the j-th layer of $M^{i,k}$. We denote the sum of the absolute values of all parameters in the j-th layer as $|\sum M_j^{i,k}|$. The contribution of the j-th layer to the overall model performance, $v(j)$, can be expressed as: $v(j) = |\sum M_j^{i,k} - \sum M_j^{i,(k-1)}|$. The larger the value of $v(j)$, the greater the impact of layer j on the performance of the final model. FL_CLIENT ranks the $v(j)$ values of all layers in the model in descending order, and selects only the parameters of the first n layers to be uploaded to the FL_SERVER for federated model aggregation. A user can set the desired value for n through FedVision. A video demonstration of the functionalities of the FedVision platform can be accessed online.[4]

Impact

FedVision has been deployment through collaboration between Extreme Vision and WeBank since May 2019. It is currently serving three large-scale corporate customers: 1) China Resources (CRC),[5] 2) GRG Banking,[6] 3) State Power Investment Corporation (SPIC).[7] CRC has business interests in consumer products, healthcare, energy services, urban construction and operation, technology and finance. It has more than 420,000 employees. FedVision has been used to help it detect multiple types of safety hazards via cameras in more than 100 factories. GRG Banking is a globally renowned AI solution provider in financial self-service industry in the world. It has more than 300,000 equipment (e.g., ATMs) deployed in over 80 countries. FedVision has been used to help it monitor suspicious transaction behaviors via cameras on the equipment. SPIC is the world's largest photovoltaic power generation company which facilities in 43 countries. FedVision has been used to help it monitor the safety of more than 10,000 square meters of photovoltaic panels.

The overview of FedVision deployment with these three customers is illustrated in Figure 7. The Federated

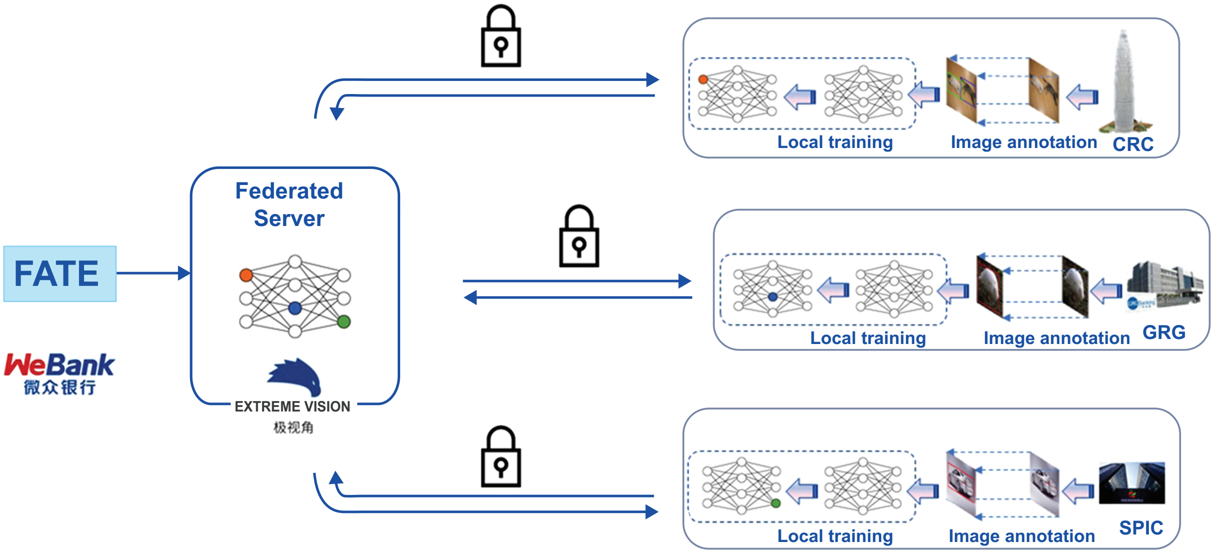


FIGURE 7 An overview of the application of the FedVision platform in the industry

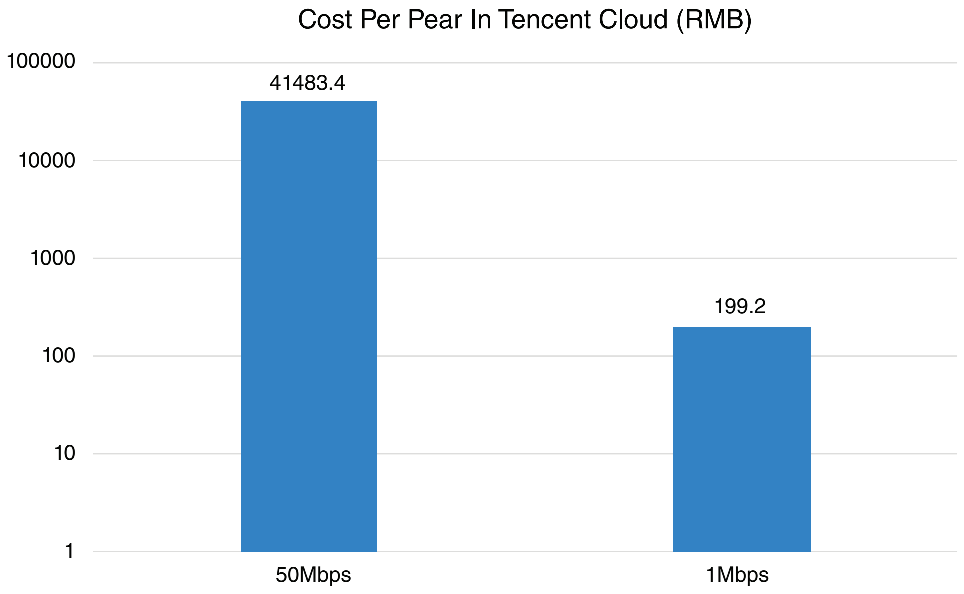


FIGURE 8 The communication cost incurred by SPIC per year. The data can be found on Tencent Cloud website in 2020

AI Technology Enabler (FATE) framework developed by WeBank is deployed through the Extreme Vision platform. It serves as the FL_SERVER. Each customer executes an FL_CLIENT under FedVision. Model parameter updates are transmitted between the FL_SERVER and the FL_CLIENTs in a secure manner. By adopting FedVision, the customers have achieved the following business improvements:

Efficiency: in the flame identification system of CRC, to improve a model, at least 1,000 sample images were needed. The entire procedure generally required five labellers for about 2 weeks, including the time of testing and packaging. As the sample images are not sufficient in practice, labeller spent a lot of time not only labeling but also waiting for the data. Thus, the total time for model optimization can be up to 30 days. In subsequent operations, the procedure would be repeated. With FedVision, the system administrator can finish labeling the images by himself. The time of model optimization is reduced by more than 20 days, saving labor cost. As data from the National Bureau of Statistics[8] show that the average annual salary for information transmission, software and information technology services is 161,352 RMB, thus each of our system save more than 10,000 RMB per operation.

Data Privacy: under FedVision, image data do not need to leave the machine with which they are collected to facilitate model training. In the case of GRGBanking, to 10,000 photos were required to train its model. Each photo is around 1 MB in size. The 10,000 photos used to require 2 to 3 days to be collected and downloaded to a central location. During this process, the data would go through two to three locations and are at risk of being exposed. With the help of FedVision, GRGBanking can leverage the local storage and computational resources at their ATM equipment to train a federated suspicious activity detection model, thereby reducing the risk of data exposure.

Cost-Saving: in the generator monitoring system of SPIC, a total of 100 channels of surveillance videos are in place in one generator facility. Under the data transmission rate of 512 KB/s for synchronous algorithm analysis and optimization, these 100 channels require at least 50 MB/sec of network bandwidth if image data need to be sent. This is expensive to implement on an industry scale. With FedVision, the network bandwidth required for model update is significantly reduced to less than 1 MB/s. This reduces communication overhead by 50 fold. Taking Tencent Cloud[9] as an example (Figure 8), it used to cost more than 40,000 RMB per year. With FedVision, SPIC just need to spend less than 200 RMB per year for this.

The improvements brought about by the FedVision platform has significantly enhanced the operations of the customers and provided them with competitive business advantages.

CONCLUSIONS

In this paper, we report on our experience addressing the challenges of building effective visual object detection models in a privacy-preserving and efficient manner through federated learning. The deployed FedVision platform is an end-to-end machine learning engineering platform for supporting easy development of FL-powered computer vision applications. The platform has been used by three large-scale corporate customers to develop computer vision-based safety hazard warning solutions in smart city applications. It has helped the customers improve their operational efficiency, achieve data privacy protection, and reduced cost significantly. To the best of our knowledge, this is the first industry application of federated learning in computer vision-based tasks.

ACKNOWLEDGMENTS

This research was supported by the R&D group of Extreme Vision Ltd, Shenzhen, China; the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-019); the RIE 2020 Advanced Manufacturing and Engineering (AME) Programmatic Fund (No. A20G8b0102), Singapore; the Joint NTU-WeBank Research Centre on Fintech (Award No: NWJ-2020-008), Nanyang Technological University, Singapore; Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR) (NSC-2019-011); and Nanyang Technological University, Nanyang Assistant Professorship (NAP). Qiang Yang is supported, in part, by China National Key Research and Development Program of China under Grant No.2018AAA0101100.

ENDNOTES

[1] https://www.webank.com/en/
[2] https://www.extremevision.com.cn/?lang=en_US
[3] https://pjreddie.com/darknet/
[4] https://youtu.be/yfiO3NnSqFM
[5] https://en.crc.com.cn/
[6] http://www.grgbanking.com/en/
[7] http://eng.spic.com.cn/
[8] http://www.stats.gov.cn/
[9] https://cloud.tencent.com/

REFERENCES

Bengio, Y., and Y. LeCun, eds. 2016. The 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings.

Bonawitz, K., H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, et al. 2019. "Towards Federated Learning at Scale: System Design." In *CoRR*, http://arxiv.org/abs/1902.01046.

Cheng, Y., D. Wang, P. Zhou, and T. Zhang. 2017. "A Survey of Model Compression and Acceleration for Deep Neural Networks." In *CoRR*, http://arxiv.org/abs/1710.09282.

Doan, A., R. Ramakrishnan, and A. Y. Halevy. 2011. "Crowdsourcing Systems on the World-Wide Web." *Communications of the ACM* 54(4): 86–96.

Kairouz, P., H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, et al. 2019. "Advances and Open Problems in Federated Learning." *CoRR* arXiv:1912.04977.

Liu, Y., A. Huang, Y. Luo, H. Huang, Y. Liu, Y. Chen, L. Feng, T. Chen, H. Yu, and Q. Yang. 2020. "FedVision: An Online Visual Object Detection Platform powered by Federated Learning." In *Proceedings of the 32nd AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-20)*, pp. 13172–9.

McMahan, H. B., E. Moore, D. Ramage, and B. A. yArcas. 2016. "Federated Learning of Deep Networks Using Model Averaging." In *CoRR*, http://arxiv.org/abs/1602.05629.

Redmon, J., and A. Farhadi. 2018. "Yolov3: An Incremental Improvement." In *CoRR*, http://arxiv.org/abs/1804.02767.

Ren, S., K. He, R. Girshick, and J. Sun. 2017. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 39(6): 1137–49.

Voigt, P., and A. v. d. Bussche. 2017. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Publishing Company, Incorporated, 1st edition.

Yang, Q., Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu. 2019. *Federated Learning*. San Rafael, CA, USA: Morgan & Claypool Publishers, p. 207.

Yu, H., C. Miao, Y. Chen, S. Fauvel, X. Li, and V. R. Lesser. 2017. “Algorithmic Management for Improving Collective Productivity in Crowdsourcing.” *Scientific Reports* 7(12541). https://doi.org/10.1038/s41598–017–12757–x.

Zhao, Z., P. Zheng, S. Xu, and X. Wu. 2018. “Object Detection with Deep Learning: A Review.” In *CoRR*, http://arxiv.org/abs/1807.05511.

AUTHOR BIOGRAPHIES

Yang Liu received the BSc. degree from Tsinghua University, in 2007, and the PhD degree from Princeton University, in 2012. She is currently the Principal Researcher of the AI Department, WeBank, China. Her research has been published in leading scientific journals, such as ACM TIST and Nature. Her research interests include machine learning, federated learning, transfer learning, multi-agent systems, statistical mechanics, and applications of these technologies in the financial industry.

Anbu Huang is a Senior Researcher at WeBank AI Department, China. Previously, he when worked as a Senior Researcher at Tencent, he led and created the world’s largest online recommendation platform for Chinese music streaming. He received his MSc degree from Tsinghua University. His research interests include deep learning, machine learning and federated learning, etc. his research topics has been published in leading international conferences.

Yun Luo is the Founder and CTO of Extreme Vision, and a PhD student of artificial intelligence at Hong Kong University of Science and Technology, under Prof Qiang Yang. She is one of source contributors to Google Tensorflow, the world’s largest open source framework for artificial intelligence, and winner of the 2017 Informatization China Artificial Intelligence Innovator competition. She was a guest speaker at the Artificial Intelligence Sub-forum of The Boao Asia Youth Forum in 2017. She holds multiple patents.

He Huang is the Co-Founder of Extreme Vision. He graduated from Shenzhen University and is proficient in Java and C++ with 8 years of experience in software development, full stack development and architecture design, and high concurrent architecture designing in IOT and AI platform.

Youzhi Liu is a Senior Product Manager in the AI department of WeBank, the world’s leading digital bank. He has been putting his effort into AI research projects and product commercialization. He used to work at Tencent Cloud as a senior product manager.

Yuanyuan Chen is a Research Engineer in the School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU), Singapore. His research interests include federated learning, and statistical learning theory, etc.

Lican Feng is a senior software development engineer and author of several software books. He mainly focuses on Cloud platform technology and cluster maintenance automation. His research interests include artificial Intelligence, inference platform, training platform and federal learning.

Tianjian Chen is the Deputy General Manager of the AI Department of WeBank, China. He is now responsible for building the Banking Intelligence Ecosystem based on Federated Learning Technology. Before joining WeBank, he was the Chief Architect of Baidu Finance, Principal Architect of Baidu. Tianjian has over 12 years of experience in large-scale distributed system design and enabling technology innovations in various application fields, such as web search engine, peer-to-peer storage, genomics, recommender system, digital banking, and machine learning.

Han Yu is a Nanyang Assistant Professor (NAP) in the School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU), Singapore. He held the prestigious Lee Kuan Yew Post-Doctoral Fellowship (LKY PDF) from 2015 to 2018. He obtained his PhD from the School of Computer Science and Engineering, NTU. His research focuses on federated learning, and algorithmic fairness. He has published over 150 research papers in book chapters, leading international conferences and journals. His research works have won multiple awards from conferences and journals.

Qiang Yang is the head of the AI department at WeBank (Chief AI Officer) and Chair Professor at the Computer Science and Engineering (CSE) Department of the Hong Kong University of Science and Technology (HKUST), where he was a former head of CSE Department and founding director of the Big Data Institute (2015-2018). His research interests include AI, machine learning, and data mining, especially in transfer learning, automated planning, federated learning, and case-based reasoning. He is a fellow of several international societies, including ACM, AAAI, IEEE, IAPR, and AAAS. He received his PhD in Computer Science in 1989 and his M.Sc. in Astrophysics in 1985, both from the University of Maryland, College Park. He obtained his BSc in Astrophysics from Peking University in 1982. He had been a faculty member at the University of Waterloo (1989-1995) and Simon Fraser University (1995-2001). He was the founding Editor-in-Chief of the ACM Transactions on Intelligent Systems and Technology (ACM TIST) and IEEE Transactions on Big Data (IEEE TBD). He served as the President of International Joint Conference on AI (IJCAI, 2017–2019) and an executive council member of Association for the Advancement of AI (AAAI, 2016–2020). Qiang Yang is a recipient of several awards, including the 2004/2005 ACM KDDCUP Championship, the ACM SIGKDD Distinguished Service Award (2017), and AAAI Innovative AI Applications Award (2016). He was the founding director of Huawei's Noah's Ark Lab (2012-2014) and a co-founder of 4Paradigm Corp, an AI platform company. He is an author of several books including Intelligent Planning (Springer), Crafting Your Research Future (Morgan & Claypool), and Constraint-based Design Recovery for Software Engineering (Springer).

How to cite this article: Liu, Y., A. Huang, Y. Luo, H. Huang, Y. Liu, Y. Chen, L. Feng, et al. 2021. "Federated Learning-Powered Visual Object Detection for Safety Monitoring." *AI Magazine* 42: 19–27. https://doi.org/10.1609/aaai.12002.