ISSN 2616-5775 Vol. 4, Issue 5: 98-103, DOI: 10.25236/AJCIS.2021.040514

Application of IOT deep learning in edge computing: a review

Houchi Li

Hunan University of Sicience and Technology, Xiangtan, 411100, Hunan, China

Abstract: *With the development of artificial intelligence technology, deep learning is widely used as a method to extract features from complex networks. However, deep learning models often run in cloud computing data centers with powerful computing capabilities. Traditional cloud computing methods rely heavily on the network, which has high latency, and has problems of security and privacy. Edge computing complements cloud computing by performing tasks at the edge of the network, resulting in significant reductions in system operation time, memory cost, and power consumption. At the same time, because it is deployed in an edge computing environment, network performance can be optimized and user privacy can be protected. This review discusses the application of deep learning on the Internet of things(IOT) in the environment of edge computing, compares the results of edge computing and cloud computing in the field of deep learning, shows the superiority of the edge computing. This paper introduces the commonly used method of edge computing, and at the same time puts forward the possible problems of edge computing in the field of deep learning.Finally, we make a prospect for the future in the cross field of edge computing and deep learning.*

Keywords: *Deep Learning; Artificial Intelligence; Edge Computing; IOT*

1. Introduction

With the rapid development of artificial intelligence, deep learning, as the most important branch of machine learning, has been widely used in academia and industry. The feature extraction of traditional machine learning mainly relies on manual extraction. For simple tasks with obvious features, manual extraction of features will appear simple and effective, but when faced with complex tasks and features are not obvious, it is often difficult for manual extraction to correctly extract features.. Deep learning is a method of automatically learning rules from complex data sets, which can discover features that are difficult to find by humans. Deep learning is a promising method to extract accurate information from sensor data of IoT devices deployed in complex environments [1-3]. For example, AlphaGo uses deep learning technology to defeat Go world champion Ke Jie many times. While showing a strong learning ability, it also provides a better vision for future AI technology [4]. Deep learning is to learn the internal laws and representation levels of samples, and has achieved outstanding results in speech and image recognition. Noda K proposed a multi-stream hidden Markov model (MSHMM) that integrates audio and visual features into a single model to perform AVSR tasks [5]. Zhang Z exemplified the single-channel and multi-channel technologies developed in the front-end and back-end of the speech recognition system, which proved the potential of deep learning in speech recognition [6]. With the advent of convolutional neural networks, deep learning has also shown a powerful ability for image recognition. Litjens G reviews the application of deep learning in image classification, target detection, segmentation, registration and other tasks [7]. As an advanced technology in machine learning, deep learning also has the advantages of strong learning ability, wide coverage, good adaptability, high data-driven upper limit, and good portability [8-9]. At the same time, there are defects that need to be improved. Among them, the most significant disadvantage is that deep learning is often accompanied by a large amount of data processing, so it requires a lot of computing power and high cost. Many calculations need to be performed in the cloud. This leads to deep learning not only poorly portable, but also in computing. The process also has a high delay [10]. At the same time, high computing power also requires high hardware requirements. CPUs often cannot meet the requirements of deep learning, and GPUs or TPUs are needed to implement them, which increases costs. The design of common deep learning models is also very complicated. A model is often stacked by multiple layers of neural networks. Innovating a model requires a lot of manpower and time. Therefore, most users can only be limited to simple applications of existing models, and the same model will have large deviations for different data training. Therefore, model selection is also an important aspect of deep learning. In addition, the training process of deep learning is often accompanied by a series of problems such as over-fitting and under-fitting.

The emergence of edge computing can well solve most of the problems encountered in deep learning. First of all, the deep learning model has good robustness with the edge computing environment. It can offload part of the learning layer at the edge, and then transfer the reduced intermediate data to the centralized cloud server [1]. Edge computing supplements cloud computing by performing certain data processing tasks at the edge of the network. Edge computing uses the characteristics of nearby computing to split the deep learning model into two parts. The more computationally intensive part is offloaded to the edge server for calculation. A small part is kept in the local terminal for calculation. Compared with traditional cloud computing, this new generation paradigm has significantly reduced the system running time, memory cost and energy consumption of deep learning applications [31]. Deep learning methods using edge computing have achieved better results in many application fields, especially in the Internet of Things industry. The auto-driving technology of automobiles needs to be equipped with various sensors, using edge computing, so that the sensors can process data closer to the vehicle, and reduce the delay of the system in the driving process. Medical wearable devices such as smart watches and trackers use edge technology, and edge gateways can provide real-time services, such as alarms or neural stimulation mechanisms. When a patient's sudden illness is detected, the shortest response time can be achieved, reducing the patient's risk [11]. This article mainly explains the application of edge computing and deep learning networks in various fields, and introduces commonly used deep learning methods and edge computing modules. Finally, I cite the problems that edge computing may encounter in the field of deep learning and a prospect for the future.

2. Background

2.1 Deep Learning Background

With the development of wireless mobile communications, the popularization of the Internet and cloud computing, the Internet of Things technology is also becoming mature. As a branch of machine learning, deep learning has already been introduced into tasks related to the Internet of Things due to its powerful ability to process data, and has achieved good results. For example, Google’s Nest Learning thermostat, which records temperature data in a structured way, and then uses algorithms to understand the user’s temperature preferences and schedule patterns [11]. Due to the high efficiency of deep learning to extract features in the complex Internet of Things environment, it has become an indispensable part of the development of the Internet of Things. The most widely used model of deep learning is the convolutional neural network, which has shown far greater capabilities than other models in terms of image classification data sets. Among them, the convolution layer extracts image features through convolution operations, the activation layer assists in expressing complex features, then the pooling layer performs feature selection and information filtering, and finally the fully connected layer connects each potential label to the pooling layer output. In addition to convolutional neural networks, deep neural networks and recurrent neural networks also have a wide range of roles in the Internet of Things. In deep learning, you need to analyze the data set first, and then select the appropriate model to get better results.

2.2 Edge Computing Background

With the penetration of the Internet of Things into our daily lives, millions of sensors and other electronic devices continue to generate data through a complex network. The traditional data processing method is placed in a cloud server for processing, but it often produces high The delay, and the user’s privacy and security cannot be guaranteed. Edge computing, as an emerging computing architecture, proposes a new solution to these problems. Edge computing places data computing at the edge of the network for processing, close to the end user, not only can significantly reduce the delay of information exchange, but also reduce the possibility of user privacy leakage [12]. For example, autonomous driving technology must provide sufficient computing power with reasonable energy consumption and have stringent real-time requirements to ensure the safety of autonomous vehicles [13]. In addition to using edge computing to bring computing closer to the end user, deep learning has shown a powerful data processing capability for highly heterogeneous data input from sensors. Therefore, the introduction of deep learning of the Internet of Things into the edge computing environment is more robust than other methods.

ISSN 2616-5775 Vol. 4, Issue 5: 98-103, DOI: 10.25236/AJCIS.2021.040514

3. Typical Applications and Architecture

3.1 Edge Computing for Autonomous Driving

With the advancement of science and technology, the Internet of Things technology has also entered a period of vigorous development. As an important part of the Internet of Things, autonomous driving technology has also received extensive attention. Car sensors transmit large amounts of data every second. If these data are uploaded to the cloud for processing, it will cause high delays, fail to respond in time, and cause serious safety hazards. Edge computing provides vehicles to complete these tasks within the specified time. At the same time, it needs to rely on deep learning technology to respond to various driving scenarios. Mohammed Al-Qizwini proposed the GoogLeNet model for autonomous driving based on the direct perception method, and compared with the CNN model, it showed even better performance. Ahmad El Sallab proposed an autonomous driving framework based on deep reinforcement learning [14]. Table 1 shows the most common deep learning models in the field of autonomous driving. Three of them use CNN. They are all used in actual needs related to image recognition such as targets and obstacles. It can be seen that As a model with powerful image feature extraction effect, CNN has been widely used in deep learning autopilot technology. In addition to choosing a suitable deep learning model, edge computing systems are also particularly important in autonomous driving. The autonomous driving edge computing system is mobile and has strict energy consumption limits. Therefore, sufficient computing power must be provided with reasonable energy consumption to ensure the safety of autonomous vehicles, even at high speeds [15]. HongjunDai considers the characteristics of autonomous driving tasks and proposes a task scheduling algorithm BRFS, and through the improved earliest deadline first algorithm, the replacement and reorganization of tasks selects a more suitable MEC server with task migration [16]. Aiming at the problems of time-varying, location-dependent and delay-constrained autonomous driving services with large content, Quan Yuan proposed a two-level edge computing architecture for autonomous driving services [17]. In response to the problem of cloud computing transmission delay, Hang Zhao proposed an edge stream data processing framework for autonomous driving, migrating computing and storage capabilities from remote cloud data centers to edge data centers [18]. Edge computing can greatly reduce the response delay of autonomous driving and improve driving safety.

3.2 Edge Computing for Medical Wearable Devices

With the improvement of the economic level, more and more people are beginning to pay attention to their own health problems. Medical wearable devices can upload the wearer's physical information in real time and respond in time, which has a large market space in society. At present, wearable devices on the market have different forms, mainly including: smart helmets, smart watches, smart glasses, etc. The edge computing server obtains a large amount of raw data through these devices for preprocessing, and then transmits it to the cloud for additional processing or long-term storage. Second, the edge gateway can provide real-time services, such as alarms or neural stimulation mechanisms, to ensure the health of the wearer [34]. In the edge computing model, two models dominate: 1) hierarchical model, 2) software-defined model [12]. The quality of the model is often evaluated by training loss, consistency (consistency of each terminal model under a decentralized architecture), privacy, data exchange cost, delay, and energy efficiency.

Epilepsy is a chronic brain disease that manifests as recurrent seizures. Globally, the incidence of epilepsy is about 1%. Patients with epilepsy are more prone to sudden death of unknown cause than normal people. Therefore, it can be timely when epileptic seizures occur. Reacting is especially important. Because the EEG of epilepsy signals is changeable, it is difficult to predict epileptic seizures, so deep learning methods are often used to find its inherent laws. YeYuan proposed a multi-view learning model that can jointly learn multi-view features from unsupervised multi-channel EEG reconstruction and supervised epilepsy detection represented by spectrograms [22]. Ramy Hussein predicts epileptic seizures through the LSTM network [23]. Table 2 lists the commonly used deep learning models for epilepsy prediction. Because epilepsy signals are one-dimensional signals, CNN and LSTM networks are often used for the feature processing of epilepsy EEG signals. Among them, when using CNN to extract the features of epilepsy EEG signals, it is necessary to first convert the one-dimensional epilepsy EEG signals into two-dimensional images to facilitate the extraction of the convolutional neural network. Mohammad-Parsa Hosseini proposed an autonomous edge computing framework for processing big data as part of the surgical candidacy decision support system [34]. Md Abu Sayeed et al. established an intelligent epilepsy detection framework at the edge of the Internet of Things, which greatly reduced the system delay compared with the cloud Internet of Things [35]. The development of autonomous detection and epilepsy localization methods will optimize epilepsy control and improve the quality of life. Epilepsy diagnosis is only a small aspect in the medical field. It can be seen that the deep learning model using edge computing architecture still has a lot of room for development in the field of medical wearable devices.

4. Summary and Challenges

This review mainly discusses the introduction of deep learning of the Internet of Things into the edge computing environment, analyzes the commonly used deep learning methods of Internet of Things devices, and demonstrates the advantages of edge computing, that is, it can make the Internet of Things devices have shorter response time and delay, and Protect the privacy of users. For tasks that require strict response time such as autonomous driving, edge computing has an unparalleled advantage over cloud computing. In the medical and health field, edge computing can also have faster response and acceptable computing power to deal with patients' serious health conditions. As an emerging technology, edge computing also has many challenges. One of the most important challenges is that it is difficult to program and manage resources and data for different applications running on different platforms and in different locations [12]. Although compared to cloud computing, edge computing can process tasks at the edge of the network, avoiding the risk of user privacy being leaked in the cloud computing center, but multi-category device access in edge computing will also bring new privacy and security issues[28].

5. Outlook

With the development of deep learning in the Internet of Things, edge computing is gradually becoming a scalable and cost-effective solution for storing and processing big data collected from a large number of biosensors. IoT devices that use deep learning can offload part of the calculations to nearby edge servers for calculations, and the energy consumption of the devices and the delay of tasks can be significantly reduced.

In addition to autonomous driving and medical wearable devices, smart cities, smart transportation, smart grids and other smart systems also interweave network communication technology with sensors, which can be regarded as a derivative of the Internet of Things technology. As more and more systems become smart Edge computing can provide these systems with less computing power and faster storage. Mobile edge computing can also be used in conjunction with 5G, which can solve the sudden increase in traffic and provide a long-term high-speed, safe Support [29]. As a new and popular technology, edge computing still has great development potential and room for development in the field of Internet of Things. It is very likely that a relatively complete edge computing technology system will be formed in the near future, which is worth exploring for researchers.

Table 1 Commonly Used Deep Learning Models for Autonomous Driving

Author	Problem	Model
[19]Aysxegul Ucxar	object detection application	LM-CNN-SVM
[32]Mohammed Al-Qizwini	Direct perception	GoogLeNet
[33]Gowdham Prabhakar	On-road obstacle detection	Faster R-CNN
[20]Guofa Li	enhancement model for low-light images	LE-net
[21]Brody Huval	car and lane detection	CNNS

Table 2 Commonly Used Deep Learning Model Framework for Epilepsy Prediction

Author	Model	Evaluate
[24]Hisham Daoud	DCAE + Bi-LSTM	Sensitivity=99.72%
[25]Spectral Power	SVM	Sensitivity=98.68%
[26]Syed Umar Amin	MCNN CCNN	Accuracy=75.7% Accuracy=73.8%
[27]Punjal Agarwal	CNN-SVM	Sensitivity=96.47±0.5%
[35]Md Abu Sayeed	DWT+Naive Bayes	Accuracy=98.65%

ISSN 2616-5775 Vol. 4, Issue 5: 98-103, DOI: 10.25236/AJCIS.2021.040514

References

[1] He L, Ota K, Dong M. Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing[J]. IEEE Network, 2018, 32(1):96-101.

[2] Voulodimos A, Doulamis N, Doulamis A, et al. Deep learning for computer vision: A brief review[J]. Computational intelligence and neuroscience, 2018, 2018.

[3] Zhao Z Q, Zheng P, Xu S, et al. Object detection with deep learning: A review[J]. IEEE transactions on neural networks and learning systems, 2019, 30(11): 3212-3232.

[4] Wang F Y, Zhang J J, Zheng X, et al. Where does AlphaGo go: From church-turing thesis to AlphaGo thesis and beyond[J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(2): 113-120.

[5] Noda K, Yamaguchi Y, Nakadai K, et al. Audio-visual speech recognition using deep learning[J]. Applied Intelligence, 2015, 42(4): 722-737.

[6] Zhang Z, Geiger J, Pohjalainen J, et al. Deep learning for environmentally robust speech recognition: An overview of recent developments[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2018, 9(5): 1-28.

[7] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis[J]. Medical image analysis, 2017, 42: 60-88.

[8] Mohammadi M, Al-Fuqaha A, Sorour S, et al. Deep learning for IoT big data and streaming analytics: A survey[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2923-2960

[9] Al-Garadi M A, Mohamed A, Al-Ali A K, et al. A survey of machine and deep learning methods for internet of things (IoT) security[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 1646-1685.

[10] Chen J, Ran X. Deep Learning With Edge Computing: A Review[J]. Proceedings of the IEEE, 2019, PP(99):1-20.

[11] Jie T, Sun D, Liu S, et al. Enabling Deep Learning on IoT Devices[J]. Computer, 2017, 50(10):92-96.

[12] Yu W, Liang F, He X, et al. A Survey on the Edge Computing for the Internet of Things[J]. IEEE Access, 2017.

[13] Liu S, Liu L, Tang J, et al. Edge Computing for Autonomous Driving: Opportunities and Challenges[J]. Proceedings of the IEEE, 2019, PP(99):1-20.

[14] Sallab A, Abdou M, Perot E, et al. Deep Reinforcement Learning framework for Autonomous Driving[J]. Electronic Imaging, 2017, 2017(19):70-76.

[15] Liu S, Liu L, Tang J, et al. Edge Computing for Autonomous Driving: Opportunities and Challenges[J]. Proceedings of the IEEE, 2019, 107(8): 1697-1716.

[16] Dai H, Zeng X, Yu Z, et al. A Scheduling Algorithm for Autonomous Driving Tasks on Mobile Edge Computing Servers[J]. Journal of Systems Architecture, 2019, 94:14-23.

[17] Yuan Q, Zhou H, Li J, et al. Toward efficient content delivery for automated driving services: An edge computing solution[J]. IEEE Network, 2018, 32(1): 80-86.

[18] Zhao H, Yao L B, Zeng Z X, et al. An edge streaming data processing framework for autonomous driving[J]. Connection Science, 2021, 33(2): 173-200.

[19] Uçar A, Demir Y, Güzeliş C. Object recognition and detection with deep learning for autonomous driving applications[J]. Simulation, 2017, 93(9): 759-769.

[20] Li G, Yang Y, Qu X, et al. A deep learning based image enhancement approach for autonomous driving at night[J]. Knowledge-Based Systems, 2021, 213: 106617.

[21] Huval B, Wang T, Tandon S, et al. An Empirical Evaluation of Deep Learning on Highway Driving[J]. Computer Science, 2015.

[22] Yuan Y, Xun G, Jia K, et al. A multi-view deep learning framework for EEG seizure detection[J]. IEEE journal of biomedical and health informatics, 2018, 23(1): 83-94.

[23] Hussein R, Palangi H, Ward R, et al. Epileptic Seizure Detection: A Deep Learning Approach[J]. arXiv, 2018.

[24] Daoud H, Bayoumi M A. Efficient Epileptic Seizure Prediction Based on Deep Learning[J]. IEEE Transactions on Biomedical Circuits and Systems, 2019, PP(99):1-1.

[25] Zhang Z, Parhi K K. Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power[J]. IEEE transactions on biomedical circuits and systems, 2015, 10(3): 693-706.

[26] Amin S U, Alsulaiman M, Muhammad G, et al. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion[J]. Future Generation Computer Systems, 2019, 101.

[27] Epileptic Seizure Prediction over EEG Data using Hybrid CNN-SVM Model with Edge Computing Services[J]. Matec Web of Conferences, 2018, 210.

ISSN 2616-5775 Vol. 4, Issue 5: 98-103, DOI: 10.25236/AJCIS.2021.040514

[28] Shi W, Jie C, Quan Z, et al. Edge Computing: Vision and Challenges[J]. Internet of Things Journal, IEEE, 2016, 3(5):637-646.
[29] Dong R, She C, Hardjawana W, et al. Deep Learning for Hybrid 5G Services in Mobile Edge Computing Systems: Learn from a Digital Twin[J]. IEEE Transactions on Wireless Communications, 2019, 18(10):4692-4707.
[30] Huh J H, Seo Y S. Understanding Edge Computing: Engineering Evolution with Artificial Intelligence[J]. IEEE Access, 2019, PP(99):1-1.
[31] Huang Y, Ma X, Fan X, et al. When deep learning meets edge computing[C]// 2017 IEEE 25th International Conference on Network Protocols (ICNP). IEEE, 2017.
[32] Al-Qizwini M, Barjasteh I, Al-Qassab H, et al. Deep learning algorithm for autonomous driving using GoogLeNet[C]// 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2017.
[33] Prabhakar G, Kailath B, Natarajan S, et al. Obstacle detection and classification using deep learning for tracking in high-speed autonomous driving[C]//2017 IEEE region 10 symposium (TENSYMP). IEEE, 2017: 1-6.
[34] Hosseini M P, Tran T X, Pompili D, et al. Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fMRI and EEG big data[C]//2017 IEEE international conference on autonomic computing (ICAC). IEEE, 2017: 83-92.
[35] Sayeed A, Mohanty S P, Kougianos E, et al. A Robust and Fast Seizure Detector for IoT Edge[C]// 2018 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS). IEEE, 2019.