

Research Article

An Intelligent Garbage Sorting System Based on Edge Computing and Visual Understanding of Social Internet of Vehicles

Xuehao Shen,¹ Yuezhong Wu ,¹ Shuhong Chen ,² and Xueming Luo³

¹College of Railway Transportation, Hunan University of Technology, Zhuzhou 412007, China

²School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China

³Scholl of Computer Science, Hunan University of Technology, Zhuzhou 412007, China

Correspondence should be addressed to Yuezhong Wu; wuyuezhong@hut.edu.cn and Shuhong Chen; shuhongchen@gzhu.edu.cn

Received 10 July 2021; Revised 6 August 2021; Accepted 21 August 2021; Published 31 August 2021

Academic Editor: Ke Gu

Copyright © 2021 Xuehao Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to enable Social Internet of Vehicles devices to achieve the purpose of intelligent and autonomous garbage classification in a public environment, while avoiding network congestion caused by a large amount of data accessing the cloud at the same time, it is therefore considered to combine mobile edge computing with Social Internet of Vehicles to give full play to mobile edge computing features of high bandwidth and low latency. At the same time, based on cutting-edge technologies such as deep learning, knowledge graph, and 5G transmission, the paper builds an intelligent garbage sorting system based on edge computing and visual understanding of Social Internet of Vehicles. First of all, for the massive multisource heterogeneous Social Internet of Vehicles big data in the public environment, different item modal data adopts different processing methods, aiming to obtain a visual understanding model. Secondly, using the 5G network, the model is deployed on the edge device and the cloud for cloud-side collaborative management, aiming to avoid the waste of edge node resources, while ensuring the data privacy of the edge node. Finally, the Social Internet of Vehicles devices is used to make intelligent decision-making on the big data of the items. First, the items are judged as garbage, and then the category is judged, and finally the task of grabbing and sorting is realized. The experimental results show that the system proposed in this paper can efficiently process the big data of Social Internet of Vehicles and make valuable intelligent decisions. At the same time, it also has a certain role in promoting the promotion of Social Internet of Vehicles devices.

1. Introduction

Social Internet of Vehicles (SIOV) is considered to be the core component of the future intelligent transportation system, and it is also one of the most promising practical technologies for 5G vertical applications [1]. Since the US Department of Transportation issued the “Intelligent Transport System (ITS) Strategic Plan” in 2015, SIOV technology has been vigorously developing around the two themes of intelligence and information sharing [2]. On November 11, 2020, the World Intelligent Connected Vehicle Conference released the “Intelligent Connected Vehicle Technology Roadmap 2.0,” which further pointed out the direction for the development of intelligent vehicle networking [3]. It is a feasible solution to use vehicle as edge

node device to provide computing services and offload tasks to the edge of the network. In recent years, with the development of social economy and the improvement of material consumption, the appearance of domestic waste has become more and more diversified and complicated, and the global waste production has also shown a cliff-like growth compared with previous years. In response to this, my country has successively introduced a series of policies. In December 2016, the fourteenth meeting of the Central Finance and Economics Leading Group hosted by General Secretary Xi Jinping proposed that “it is necessary to accelerate the establishment of a garbage disposal system for classified release, classified collection, classified transportation, and classified treatment, and form a waste treatment system based on the rule of law, promoted by the

government, and a garbage classification system with participation of the whole people, urban and rural planning, and local conditions, and strive to increase the coverage of the garbage classification system.” The latest revision of the “Law of the People’s Republic of China on the Prevention and Control of Environmental Pollution by Fixed Wastes” in 2020 requires that local people’s governments at or above the county level should speed up the establishment of a domestic waste management system for classified release, classified recycling, transportation, and treatment. At present, China’s economy is developing rapidly, people’s quality of life has been greatly improved, public environmental issues have become the focus of attention, and one of the key factors affecting the public environment is the garbage issue.

At this stage, garbage classification is mainly concentrated in the outdoor public environment for fixed groups of people to deal with according to categories. There are problems such as high labor intensity, low sorting efficiency, and poor working environment, people’s low awareness of classification, and a wide variety of garbage. Therefore, from the perspectives of practicability, environmental protection, and intelligence, it is of great significance to study and design an intelligent garbage sorting vehicle system based on edge computing of SIoV. In recent years, with the rapid development of artificial intelligence and robotics, service robots have attracted widespread attention. At present, there have not been public reports about the work carried out by service robots for autonomous garbage detection and classification. At the same time, service robots are one of the edge devices in SIoV. Therefore, it is of great practical significance to implement garbage classification and detection algorithms on service robots. However, only using the detection and classification model can only realize the identification and positioning of garbage, and the degree of intelligence is not high. If you want to make robots have the ability to recognize and discriminate objects in a public environment like humans, for example, humans can understand what they see. The items in the scene can be inferred and classified based on the association and imagination based on these items, so they should not only rely on the appearance and geometric characteristics of the items, but also rely on the guidance and reasoning of the high-level prior knowledge of the items. Public environmental goods information has the characteristics of diversity, semantics, and relevance, so it can be considered to use knowledge graphs to express and store this rich prior knowledge in a structured form. Then, the effective garbage detection and classification algorithm is used to complete the identification and positioning of items. The use of 5G networks to realize collaborative computing between edge device nodes and the cloud and intelligent decision-making of garbage classification on the big data of items in the scene are the key issues studied in this paper. The main innovations and contributions of this paper include the following: First, build a new visual understanding model. This model uses the knowledge graph to uniformly characterize and store the multimodal information of items in the public environment and combines the YOLOv4 detection algorithm to identify and locate items in the scene; the second is to propose the use of cloud-side collaborative

computing. Using the 5G network, the visual understanding model deployed on the edge device and the cloud are used for cloud-side collaborative management. The cloud is used to store a large amount of data, while avoiding the waste of edge node resources, while ensuring the data privacy of the edge node; the third is to build an intelligent garbage sorting system based on edge computing and visual understanding of SIoV. Common items can be detected, identified, and classified by edge devices; abnormal items can be connected to the cloud for identification, reasoning, and decision-making.

The organizational structure of the rest of this paper is as follows. Related work will be discussed in the second part. The third part introduces the overall design of the system in detail. The fourth part mainly introduces the experimental setup and result analysis. The fifth part is the summary of this paper.

2. Related Work

2.1. Deep Learning. Deep learning is an important breakthrough in the field of artificial intelligence in the past decade, and it is widely used in target detection and classification. Target detection can be divided into two categories in terms of methods. One is the regression-based one-stage algorithm represented by the YOLO [4] series and the SSD series, and the other is based on the candidate region two-stage algorithm represented by Fast R-CNN and Faster R-CNN. In recent years, many researchers have used deep learning technology in the research of garbage identification, classification, and detection. Literature [5] proposed a method of automatic identification of garbage types combined with ResNet-50 and multilevel SVM. GCNet [6] consists of a feature extractor and a classifier. The feature extractor uses ResNet101 as the backbone network, which contains 5 Bottlenecks, and each Bottleneck adds an attention mechanism, and then the extracted different features are merged. Literature [7] proposed an improved network’s robustness framework DNN-TC for junk image recognition. The network adds two fully connected layers after the output layer and the global average pooling layer of the classification latitude to reduce model parameter redundancy. The last layer uses the log softmax function to calculate the confidence of each label. Literature [8] uses DenseNet with higher detection accuracy and at the same time modifies the connection mode between dense blocks to achieve the purpose of improving the detection speed. Ma Wen et al. [9] used ResNet50 to replace the original VGG16 basic network in Faster R-CNN and Soft-NMS instead of the original NMS, which improved accuracy and reduced time. Wang Mingjie [10] used K-mean++ to determine the size of the prior frame and then used migration learning to complete the location and classification of garbage using YOLOv3 and classified the waste into recyclable items, dry garbage, wet garbage, and hazardous garbage. However, with the increasing demand for garbage classification by mobile edge devices and considering the accuracy and real-time performance, this paper uses YOLOv4 as the basic network. However, in order to distinguish it from the existing YOLO

detection algorithm, the model not only relies on a large amount of labeled data to train and fit a large number of parameters for prediction, but also considers the role of prior knowledge to guide the reasoning of the model. Therefore, it is planned to add a knowledge graph on the basis of the YOLOv4 algorithm to further enhance the intelligent level of the system.

2.2. Knowledge Graph. The knowledge graph can use the knowledge triples composed of nodes and relationships to intuitively represent the association relationships between items in the scene, and it can be stored in a structured form at the same time. Therefore, in a public environment, it is a very effective method to use knowledge graph to express rich visual associate on information and prior knowledge of objects. Knowledge Graph (KG) [11] is a technical method that uses graph models to describe the relationship between knowledge and modeling the world. KG was first applied to improve the capabilities of search engines; after that, KG showed greater application value in assisting intelligent question and answering, natural language processing, big data analysis, recommendation calculations, and interpretable artificial intelligence [12–14]. Marino et al. [15] studied the application of structured prior knowledge in the form of knowledge graphs in image classification. Jiang et al. [16] proposed a hybrid knowledge routing module in view of the current detection algorithm ignoring the semantic association information of the target in the scene and the long tail phenomenon of the sample size distribution of different categories. Chen et al. [17] introduced statistical target objects and their possible coexistence of prior knowledge to constrain the relationship prediction space to improve the accuracy of the model in fewer categories. Wang et al. [18] introduced the prior knowledge of the association between the characters in the scene and the surrounding objects and performed explicit reasoning based on knowledge. Wu et al. [19] proposed a visual question and answer method, which constructs a textual representation of the semantic content of an image and merges it with the textual information from the knowledge base to achieve a deeper understanding of the scene.

2.3. Edge Computing. As a new paradigm, edge computing can sink the computing functions and services of the cloud to the network edge devices, providing real-time data analysis and intelligent processing nearby, which can effectively solve the problems of network congestion and network delay caused by the transmission and processing of massive data. At present, edge computing accelerates the transformation and upgrading of the economy by providing key-capabilities such as computing, network, and intelligence nearby. It has gradually become a new direction of the computer system and a new format in the information field and has received extensive attention from academia and industry. With the popularization and development of products and application scenarios such as smart phones, smart homes, and smart connected cars, artificial intelligence is gradually migrating from the cloud to the embedded

end of the edge, and intelligent edge computing has emerged from this [20]. The concept based SIOV is an extension of the Internet of Things. Real-time collection of vehicle operating data is achieved through on board sensing units, roadside acquisition modules, vehicle-to-road communication units, and other equipment, then builds a data platform for monitoring large-scale vehicle real-time operating information, and provides various data service [21]. In the era of the Internet of everything, the massive amounts of data generated by various smart devices have put forward higher requirements on computing, storage, and network service capabilities. In order to relieve the computing pressure on the cloud and at the same time improve the computing power and operating efficiency of the mobile side, some services are deployed at the network edge close to the mobile side to build a mobile edge computing system [22–24]. The introduction of edge computing based SIOV is an inevitable trend. However, edge computing is deployed near the network infrastructure. On the one hand, it is vulnerable to attacks from edge vehicles and network infrastructure such as counterfeiting, privacy theft, and false information; on the other hand, unauthorized internal attackers may also access and steal storage at the edge. Sensitive information is in the data center [25–27]. Therefore, the efficient processing and valuable intelligent decision-making [28] based SIOV big data on the edge device side can protect the privacy and safety of vehicles in the edge computing of SIOV.

2.4. SLAM. Simultaneous localization and mapping (SLAM) is a process in which the robot uses its own vision, laser, and other sensors to complete its own positioning while constructing environmental maps and path planning [29]. The SLAM system that uses the camera to collect image information as the source of environmental perception information is called Visual SLAM (VSLAM). Compared with other SLAM systems, VSLAM can perceive richer colors, textures, and other environmental information. With the rapid development of deep learning technology, it has very successful applications in various fields of Computer Vision (CV). For example, SLAM technology is playing an increasingly important role in the fields of service robots and driverless cars. In this context, in recent years, more and more SLAM researchers use deep learning-based methods to extract environmental semantic information to obtain high-level scene perception and understanding and apply it in the VSLAM [30] system to assist VSLAM. The system improves positioning performance and map visualization, thereby giving robots more efficient human-computer interaction capabilities. The combination of deep learning and SLAM improves the limitations caused by manual design features and potentially improves the learning ability and intelligent level of the robot [31]. VSLAM can construct a 3D map of the surrounding environment and calculate the position and direction of the camera. The combination of deep learning and SLAM is a hot research direction in recent years. Among them, the semantic SLAM combining VSLAM and deep learning obtains environmental geometric information during the mapping process and at the same time recognizes

independent objects in the environment and obtains semantic information such as their positions, poses, and individual contours, which expands the research content of traditional SLAM problems. Integrate some semantic information into SLAM research to cope with the requirements of complex scenarios [32]. Therefore, this paper applies SLAM on the edge device to further enhance the autonomy of the system.

3. Overall System Design

3.1. Overall Design. This paper experimentally designs an intelligent garbage sorting system based on edge computing and visual understanding of SIOV. The overall architecture of the system is shown in Figure 1. This architecture diagram is divided into three parts:

- (1) Visual understanding model: First, the KG is used to uniformly characterize and store the multimodal information knowledge of items in the public environment. The YOLOv4 detection algorithm is used to identify and locate the items in the scene, and the constructed multimodal KG is combined with the YOLOv4 visual detection method. Construct a visual understanding model, as shown in Figure 2. Then put the model on the cloud server for large-scale data training and then deploy it to the NVIDIA Jetson Nano development board after the training is completed. Finally, when the edge device is in a public scene, use the camera to collect pictures in real time and transmit the pictures wirelessly to the development board. The development board uses the trained model to detect whether there is garbage in the picture, and if it exists, it sends it to the driver board STM32 through the serial port. STM32 controls the robotic arm to grab the garbage and put it into the corresponding garbage bin according to the recognition and classification results. The judgment of this process is to realize the intelligent decision-making of items through intelligent question and answer technology and realize the goal of garbage classification.
- (2) Cloud edge collaboration: What is used here is an edge-oriented cloud-side collaborative computing form. In this form, the cloud is only responsible for the initial training work, and the model is downloaded to the edge after the training is completed. While performing computing tasks at the edge, it will also use real-time on-site data to perform subsequent calculations on the model. This model can meet individual application requirements, make better use of local data, and avoid waste of edge node resources to ensure the edge data privacy of the node.
- (3) Edge device applications: The mapping navigation unit is based on the ROS distributed framework, uses lidar to collect the environmental information of the cleaning area, realizes the SLAM function based on the scanning matching algorithm, and uses the optimal path algorithm to autonomously plan and

traverse the cleaning area. During the traversal process of the edge sorting device, the target detection algorithm detects and classifies the real-time images obtained by the camera and obtains the coordinates and angle information of the target as the input information of the sorting control unit and controls the robotic arm to perform the garbage capture task.

In order to enable edge device to achieve intelligent and automated garbage classification in a public environment, while avoiding network congestion caused by simultaneous access to a large amount of data, and considering the combination of mobile edge computing and SIOV, in order to give full play to the high bandwidth and low latency characteristics of mobile edge computing, this paper builds an intelligent garbage sorting system based on edge computing and visual understanding of SIOV (Figure 2). First of all, according to the existence of two modalities of video and image in the public environment, the YOLOv4 detection algorithm is used to extract the location of the entity category; use BLSTM-LCRF and PCNN-BLSTM-Attention proposed by Wang Huan et al. [33] to extract entities and relationships from text modalities. The open source structured data collected from the Internet and the entity relationship extracted above form a knowledge triple. Secondly, the knowledge triples are used to uniformly represent and store the semantic description information, attribute information, and spatial location information of the items in the scene using the KG. Finally, when detecting and classifying garbage items in a public environment, the YOLOv4 detection algorithm will perform real-time detection to obtain its location and category information. At the same time, it will use OCR technology to obtain the description information of the item's outer packaging, and the previously constructed KG will be further developed through VQA technology. The auxiliary detection model matches and determines whether the item is garbage and what type of garbage in the form of intelligent question and answer and makes further intelligent decision-making.

3.2. Multimodal Knowledge Graph. With the continuous popularization of the Internet and media technologies, information from different sources such as text, images, video, audio, etc. collectively portrays the same or related content, presenting complex and multilevel semantic relationships, forming "multimodal" information. First of all, multimedia data containing different modalities present internally synchronized semantic associations, while information from different sources and modalities across media presents dynamic, complex, multilayered temporal and semantic associations. Secondly, the cross-media forms are heterogeneous, the content is diverse, and the distribution is complex. The traditional analysis and processing methods are mostly based on the assumption of independent and identical distribution, and it is difficult to effectively utilize and learn the massive and complex cross-media information. Finally, the application scenarios involved in cross-media are more extensive, such as cross-media content

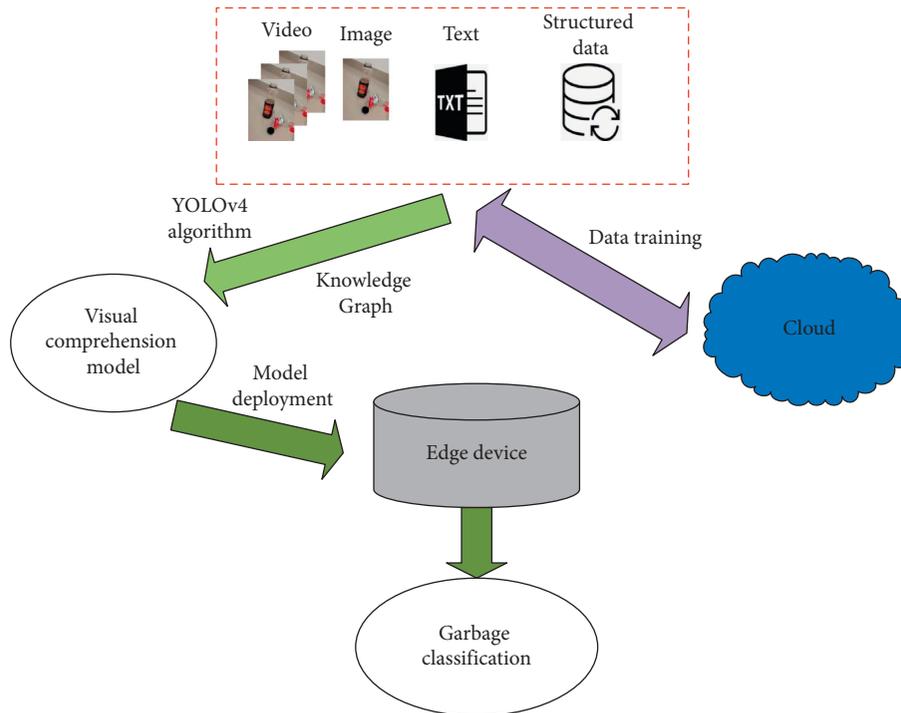


FIGURE 1: System overall architecture diagram.

search, recommendation, question and answer, etc. Abundant item data has been accumulated in public scenarios, such as “item name,” “item packaging,” “item category,” and “item production information.” However, the data are not related to each other, failing to form effective knowledge. Constructing the item data into a KG and developing upper-level applications can effectively enable garbage classification in public scenarios, solve the pain points of the scenario, and give full play to the value of knowledge. The construction and application process of the KG is shown in Figure 3.

The process of building knowledge graphs of objects in public scenarios: ontology design, knowledge extraction, knowledge mapping, knowledge fusion, and disambiguation. Ontology design is a process of knowledge modeling. This process requires the realization of ontology design by summarizing the knowledge in the field. Generally speaking, RDFS, OWL, and other languages can be used for modeling. The semiautomated form realizes the paper knowledge modeling in public scenarios and quickly completes the paper KG ontology design. The ontology includes concepts such as item names, item attributes, and associations between items. Knowledge extraction completes the extraction of knowledge triples based on relevant algorithms by collecting relevant data. For structured data, after simple conversion, triples can be generated. For unstructured data, the document format needs to be converted first to obtain easy-to-handle text formats such as txt and docx; literature [30] proposed BLSTM-LCRF and PCNN-BLSTM-Attention models to extract entities and relationships from text data. Knowledge mapping, the triples obtained in the information extraction stage, needs to map the extraction results to the ontology through knowledge mapping, in order to generate

a KG. If the amount of data is large, the process can be accelerated with the help of big data technologies such as Hadoop. Finally, it is necessary to integrate and disambiguate heterogeneous data under a unified standard for the knowledge from different data sources to complete the creation of the KG. For example, an item with the same apple name can actually refer to both fruits and apple mobile phones. Typical applications included in the paper KG based on public scenarios include intelligent question answering, intelligent reasoning, and intelligent decision-making.

3.3. YOLOv4 Network. Based on the YOLOv3 algorithm, Alexey B officially launched YOLOv4 in 2020 (shown in Figure 4). Its network structure is divided into input, backbone feature extraction network (CSPDarknet53), enhanced feature extraction network (Spatial Pyramid Pooling, SPP, and Path Aggregation Network, PANet), and detection network (YOLO Head). The principle of YOLOv4 model detection is to divide the image into an $S \times S$ grid, where each cell in the grid is responsible for predicting B location bounding boxes and conditional probabilities belonging to C categories, and finally output whether the bounding box contains the target and the bounding box confidence information. YOLOv4 uses CSPDarknet53 instead of Darknet53 in its backbone network. The main changes are as follows: first, the introduction of CSPNet (Cross Stage Partial Network) network structure. The gradient variability is integrated and mapped to the feature map from beginning to end, and then the feature map is divided into two parts: one part performs residual stacking operation, and the other part is directly combined with the last convolution result, which effectively solves the gradient information factor in

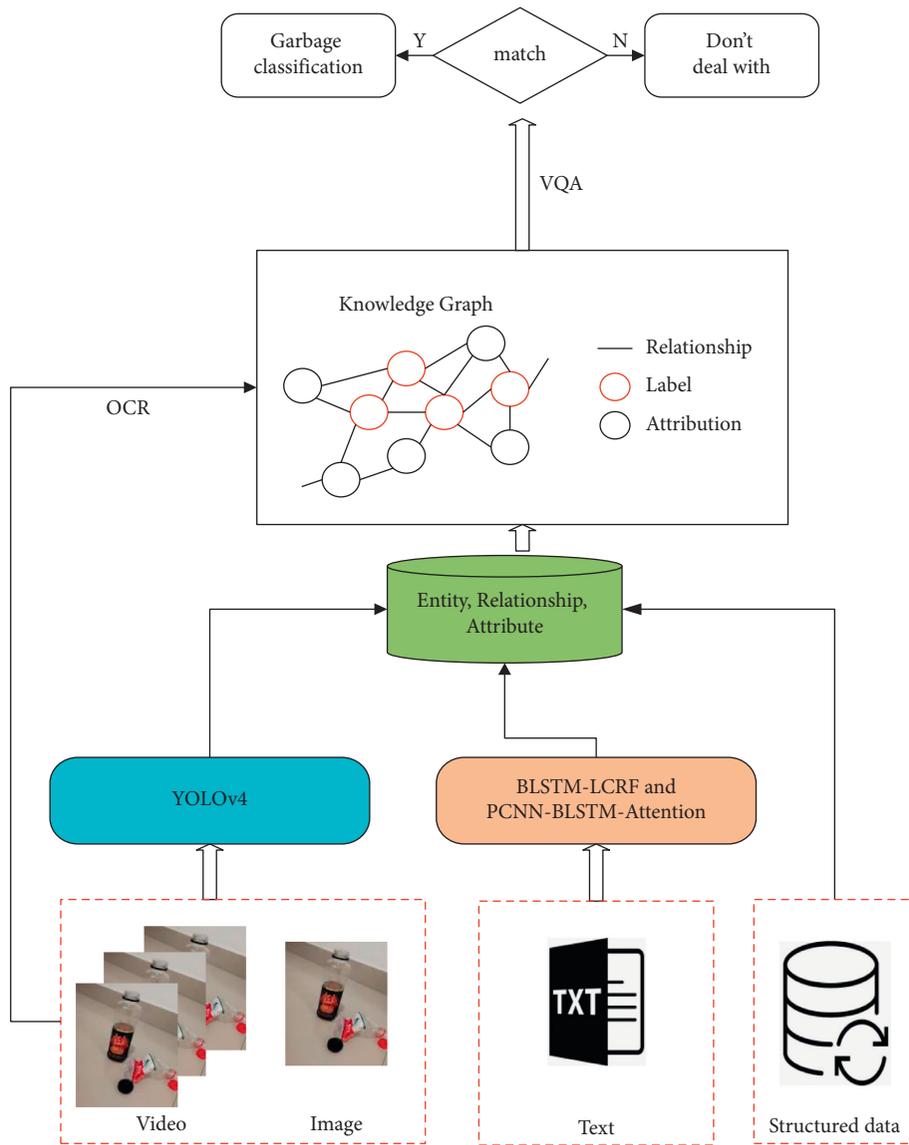


FIGURE 2: Visual comprehension model.

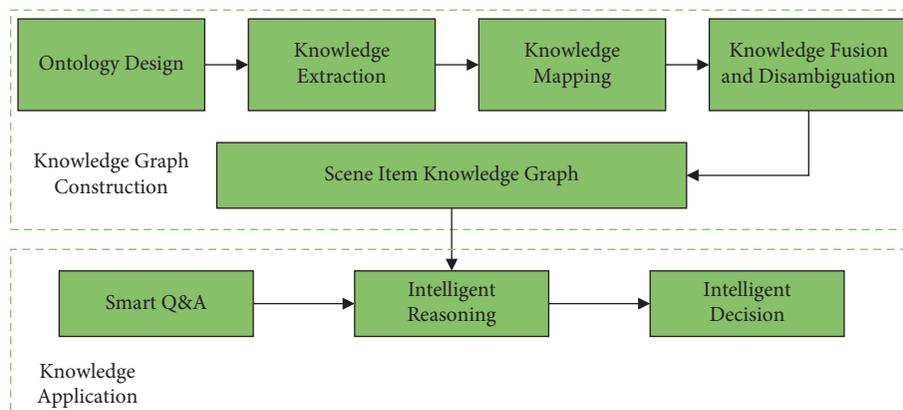


FIGURE 3: Knowledge graph construction and application flow chart.

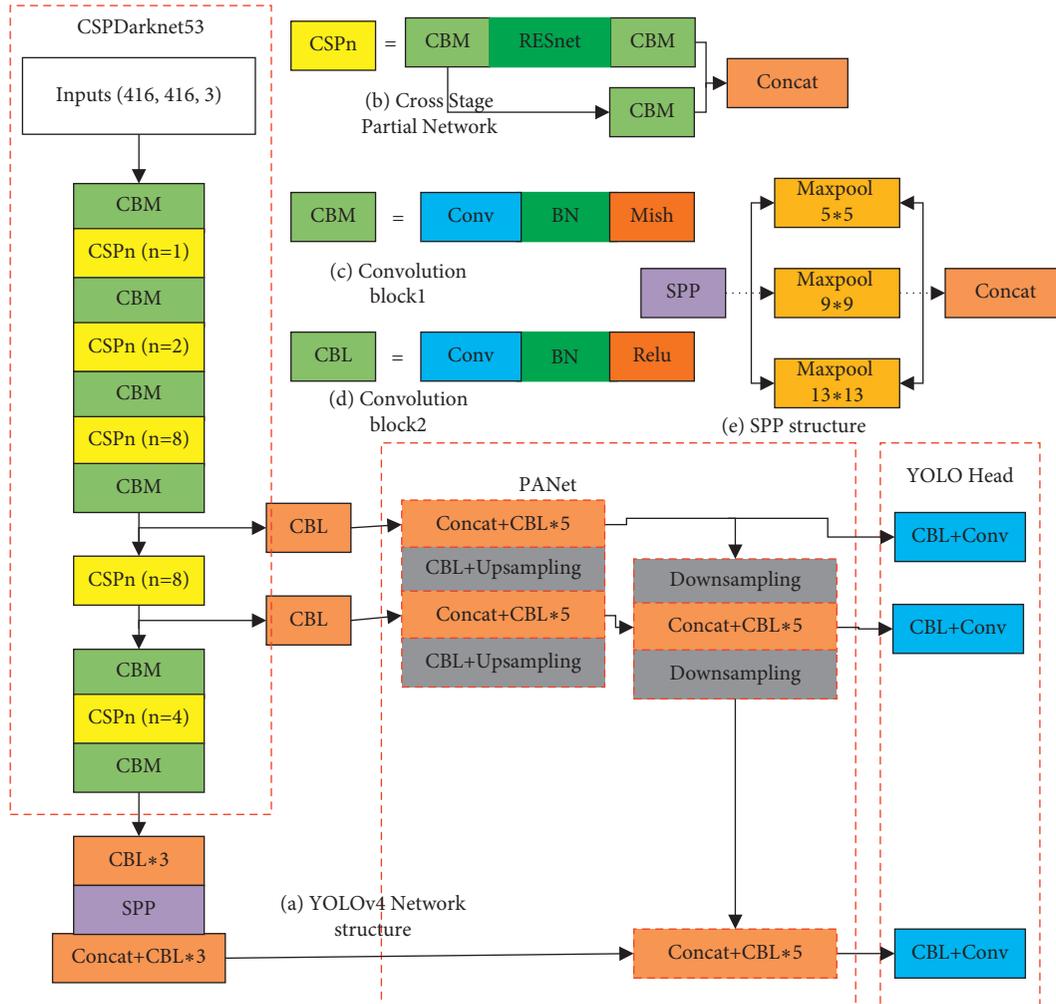


FIGURE 4: YOLOv4 network structure diagram.

the deep network processing; repeated learning causes the problem of increased calculation. The second is to replace the activation function LeakyReLU with the Mish activation function, so that the overall network detection has a higher accuracy rate. Introduce the Spatial Pyramid Pooling (SPP) layer and use the maximum pooling of different pooling core sizes to pool the input feature layer, using the maximum pooling method of 1×1 , 5×5 , 9×9 , and 13×13 , respectively. This structure can use different dimensions of the same image as input to obtain pooling features of the same length. The SPP network can effectively increase the range of feature acceptance and extract more important contextual features without reducing the network operation speed. PANet is created on the basis of FPN (Feature Pyramid Networks) in YOLOv3. It is a bottom-up path enhancement designed to promote information flow and use accurate low-level positioning information to enhance the overall feature level, thereby shortening the information path between the low-level and top-level features. This structure makes full use of feature fusion and changes the previous additive fusion method to multiplicative fusion, so that the network has higher detection accuracy. The detection part still uses YOLO Head detection using the YOLOv3 algorithm.

4. Experimental Setup and Result Analysis

4.1. Experimental Environment Configuration. The experiment in this paper is completed under the Ubuntu system. CUDA is a general parallel computing architecture launched by NVIDIA; CUDNN is a GPU acceleration library for deep neural networks. The data is trained through the cooperation of the two, and the experimental environment is configured, as shown in Table 1.

First, the collected image data is cleaned to remove some invalid images; then the processed valid data set is labeled according to the PASCAL VOC data set format; then the migration learning method is adopted, and the pretraining of yolov4.pt provided on the official website is used. The weight is used as the initial parameter of network training; finally, after 70 epochs of iteration, the loss function reaches a minimum and tends to a balanced state, and the training is stopped to obtain the best weight file. The model training loss function curve in this paper is shown in Figure 5. Some network parameter descriptions are shown in Table 2.

The model training process in this paper: calculate and predict the input data through forward propagation to obtain the loss function of the network model, then use the

TABLE 1: Experiment environment configuration.

Project	Experimental environment
System	Ubuntu18.04
Programming environment	Pycharm
GPU	NVIDIA TITAN RTX
Memory	24 GB
Pytorch version	Pytorch1.6
Python version	Python3.6
CUDA version	CUDA10.1
CUDNN version	CUDNN7.6
Data bases	Neo4j4.2.2
Java runtime environment	JDK15.0.1

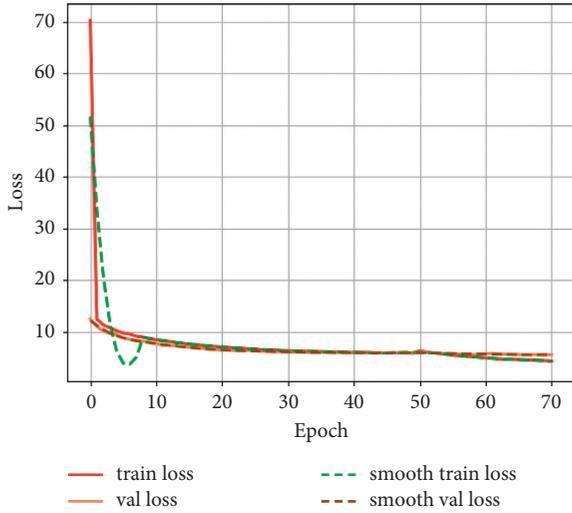


FIGURE 5: Loss curve of our model.

TABLE 2: Network parameter description table.

Parameter name	Parameter values
Learning rate	0.001
Momentum	0.9
Decay	0.0005
Batch size	64

combination of backpropagation and gradient descent to find the direction of gradient descent, and update the model weight parameters, and the whole process is repeated iteratively, and a network model with better detection effect is finally obtained. The loss function used in this network model is as follows:

$$\begin{aligned}
\text{Loss} = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B W_{ij}^{\text{obj}} [(x_i - x'_i)^2 + (y_i - y'_i)^2] \\
& + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B W_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{w'_i})^2 + (\sqrt{h_i} - \sqrt{h'_i})^2 \right] \\
& + \sum_{i=0}^{S^2} \sum_{j=0}^B W_{ij}^{\text{obj}} (C_i - C'_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B W_{ij}^{\text{noobj}} (C_i - C'_i)^2 \\
& + \sum_{i=0}^{S^2} W_{ij}^{\text{obj}} \sum_{c \in \text{classes}} [(p_i(c) - p'_i(c))]^2.
\end{aligned} \tag{1}$$

It can be seen from formula (1) that the entire loss function is composed of three parts: the first two lines in the formula represent the error of the center coordinates and width and height of the prediction box, which is the first part, where (x_i, y_i, w_i, h_i) represents the center coordinates and width and height of the prediction box, which means marking the center coordinates and width and height of the box. The third row represents the confidence error, which is the second part. The left half of the third row represents the confidence error that the prediction box contains the target, and the right half represents the confidence error that the prediction box does not contain the target, where C_i indicates that the prediction box contains the object confidence, where C'_i represents the IOU of the predicted box and the marked box. The fourth row represents the category error of the target, which is the third part, where $p_i(c)$ represents the marked category value and $p'_i(c)$ represents predicted category value. In formula (1), W_{ij}^{obj} means that the target is detected by the j prediction frame in the first i grid, and W_{ij}^{noobj} means that the target is not detected, and W_i^{obj} means that the target is in the first i grid.

4.2. Data Collection. The data set used in this paper has a total of 15,000 domestic garbage pictures, most of which come from the data set in the garbage classification competition held by Alibaba Cloud Tianchi and some pictures of domestic garbage collected by the author. The data set can be divided into four categories in general, namely, recyclable garbage, kitchen waste, hazardous garbage, and other garbage. Each category contains multiple objects. Among them are recyclable trash: power bank, bag, wash supplies, plastic toy, plastic utensils, plastic hangers, glassware, metalware, courier bags, plug wire, old clothes, ring-pull can, pillow, plush toy, shoes, cutting board, carton, wine bottle, metal food can, ironware, wok, edible oil drum, drink bottle, paper books; harmful trash: dry battery, unguentum, expired drugs; other trash: disposable snack box, stained plastic, butt, toothpick, flowerpot, chinaware, chopsticks, stained paper; use the LabelImage tool to label the items in the picture, and divide the data set into a training set and a test set at a ratio of 8 : 2.

4.3. Experimental Results and Analysis. The evaluation criteria of the results of this experiment are mainly Precision (P), Recall (R), Mean Average Precision (MAP), and detection speed Frames Per Second (FPS). Among them, P represents the ratio of the real samples in the recognized positive samples, namely,

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{2}$$

Among them, R represents the proportion of correctly identified positive samples in the total number of samples, namely,

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

TABLE 3: Algorithm detection result.

Algorithm	P (%)	R (%)	MAP (%)	FPS
YOLOv4	84.3	81.9	72.5	26
Our model	84.6	82.3	73.8	24

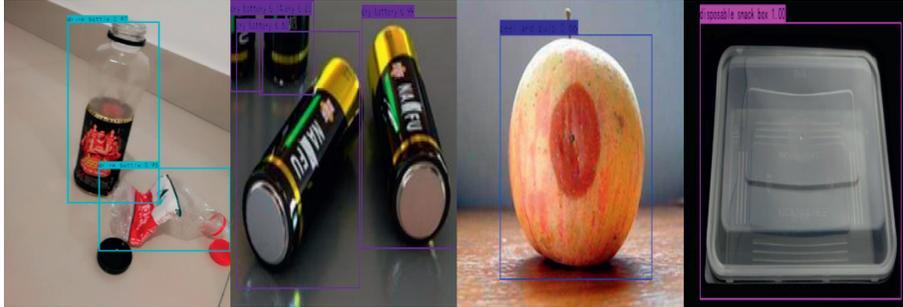


FIGURE 6: Visualization result graph.

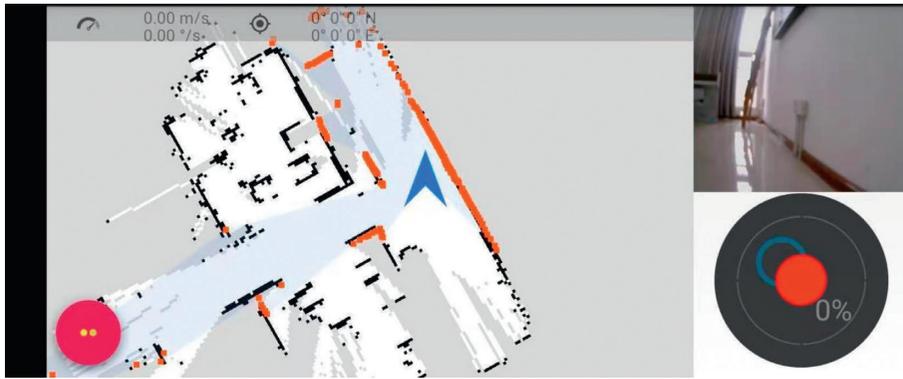


FIGURE 7: Build map.



FIGURE 8: Navigation.

TP is the number of positive samples that correctly classify the target, FP is the number of positive samples that are incorrectly classified as the target, and FN refers to the number of positive samples that are incorrectly classified as negative samples.

In order to verify the effectiveness of this test method, the visual understanding model and YOLOv4 proposed in this paper were trained and verified with different algorithms according to the network parameters in Table 2. The P , R , MAP , and FPS are shown in Table 3.



FIGURE 9: Recognition result.

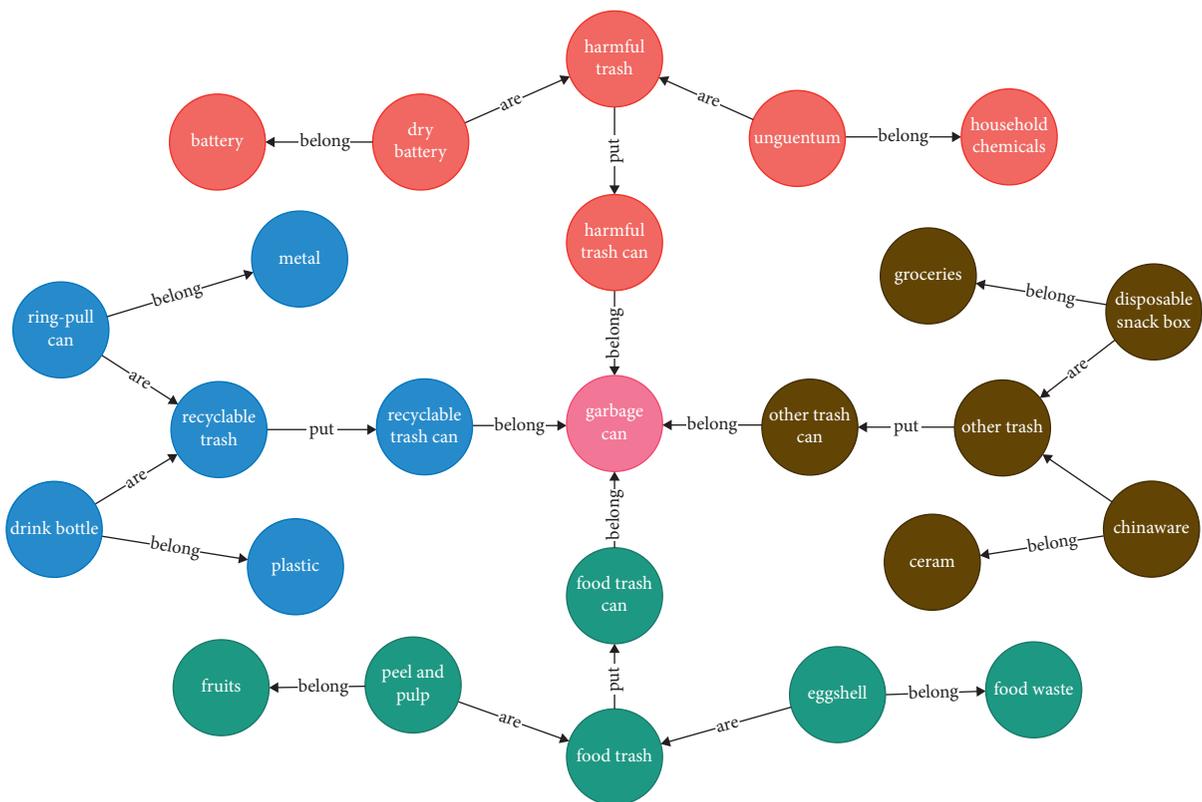


FIGURE 10: Local knowledge graph.

From the results in Table 3, the model proposed in this paper has better performance than YOLOv4. Under the premise of equivalent detection speed, the MAP can reach nearly 74%. Figure 6 shows part of the visualization results of the model detection proposed in this paper.

Taking into account the limited equipment in the actual public environment, this paper uses a trolley to simulate an

indoor scene. The results of mapping, navigation, and recognition results are shown in Figure 7, Figure 8 and Figure 9.

Figure 10 shows the constructed knowledge graph of the local scene. Different colors represent different types, among which red is harmful trash, blue is recyclable trash, green is food trash, and gray is other trash. Through the YOLOv4 algorithm, the entity name and location information of the



FIGURE 11: Classification result.

item in the scene can be obtained, the identified entity target object is matched with the entity in the KG, and the location relationship or attribute information between the entities is used to make further intelligent decisions and judgments. If the item is garbage, determine the specific type of garbage and return it to the corresponding garbage bin to facilitate the next step of sorting by the edge device; if it is not garbage, the model will not process it. Figure 11 is the result of querying the target entity in the KG and performing garbage classification. For example, a deformed beverage bottle on the ground is a recyclable trash and should be placed in a recyclable trash can.

5. Conclusions

This paper builds an intelligent garbage sorting system based on edge computing and visual understanding of SIoV. Experimental results show that the system can provide efficient and valuable intelligent decision-making, free human hands, reduce back-end waste processing, and improve work efficiency; at the same time, through cloud-side collaborative computing, the use of edge devices can be fully utilized, which can avoid cloud network congestion. It also guarantees the data privacy of edge nodes. In addition, due to the limited data set currently collected, the constructed KG is not complete, and the classification effect of some

uncommon or severely defaced items is relatively poor. On the one hand, future research work can add multimodal data analysis experiments to visual understanding; on the other hand, it can also consider using 5G networks to achieve collaborative work between multiple edge devices and increase related algorithm analysis and experiments in SIoV.

Data Availability

The data used to support the findings of this study are not applicable because the data interface cannot provide external access temporarily.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant nos. 2019YFE0122600 and 2018YFB1700200, in part by the Hunan Provincial Key Research and Development Project of China under Grant no. 2019GK2133, in part by the Natural Science Foundation of Hunan Province under Grant no. 2021JJ50050, and in part

by the Scientific Research Project of Hunan Provincial Department of Education under Grant no. 19B147.

References

- [1] Y. Xiao, H. Liu, and X. H. Cheng, "Key technologies of internet of vehicles and their development trends and challenges," *Communications Technology*, vol. 54, no. 1, pp. 1–8, 2021.
- [2] Y. Wei, "Analysis on the development of internet of vehicles based on 5G communication technology," *Application of Mechanics-electronics Technology*, vol. 54, no. 3, pp. 223–225, 2019.
- [3] F. Gao, "Interpretation of intelligent connected vehicle technology Roadmap 2.0," *Internet of Things Technology*, vol. 11, no. 1, pp. 3–4, 2020.
- [4] A. Bochkovski and C. Y. Wang, "YOLOv4: optimal speed and accuracy of object detection," 2020, <https://arxiv.org/abs/2004.10934v1>.
- [5] O. Adedeji and Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network," *Procedia Manufacturing*, vol. 35, pp. 607–612, 2019.
- [6] Z. Y. Dong and W. G. Han, "Classification algorithm of garbage images based on convolutional neural network," *Computer Systems & Applications*, vol. 29, no. 8, pp. 199–204, 2020.
- [7] A. H. Vo, L. Hoang Son, M. T. Vo, and T. Le, "A novel framework for trash classification using deep transfer learning," *IEEE Access*, vol. 7, pp. 178631–178639, 2019.
- [8] C. Bircanoglu, M. Atay, F. Beser, M. A. Kizrak, and O. Genc, "RecycleNet: intelligent waste sorting using deep neural networks," in *Proceedings of the 2018 IEEE International Conference on INnovations in Intelligent Systems and Applications (INISTA)*, July 2018.
- [9] W. Ma, J. Yu, and J. Y. Chen, "Garbage detection and classification based on improved Faster R-CNN," *Computer Engineering*, vol. 47, no. 8, pp. 294–300, 2020.
- [10] M. J. Wang, "Automatic garbage location and classification based on YOLOV3," *Wireless Internet Technology*, vol. 16, no. 20, pp. 110–112, 2019.
- [11] A. Singhal, "Official Google Blog: Introducing the Knowledge Graph: things, not strings," 2012, <http://googleblog.blogspot.pt/2012/05/introducing-knowledge-graph-things-not.html>.
- [12] M. Li, Z. Sun, S. Zhang, and W. Zhang, "Enhancing knowledge graph embedding with relational constraints," *Neurocomputing*, vol. 429, pp. 77–88, 2021.
- [13] Z. Li, H. Liu, Z. Zhang, T. Liu, and J. Shu, "Recalibration convolutional networks for learning interaction knowledge graph embedding," *Neurocomputing*, vol. 427, pp. 118–130, 2021.
- [14] F. Gong, M. Wang, H. Wang, S. Wang, and M. Liu, "SMR: medical knowledge graph embedding for safe medicine recommendation," *Big Data Research*, vol. 23, 2021.
- [15] K. Marino, R. Salakhutdinov, and A. Gupta, "The more you know: using knowledge graphs for image classification," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20–28, 2017.
- [16] C. Jiang, H. Xu, X. Liang, and L. Lin, "Hybrid knowledge routed modules for large-scale object detection," in *Proceedings of the In Advances in Neural Information Processing Systems*, pp. 1552–1563, Montreal Convention Centre, Montreal, Canada, December 2018.
- [17] T. Chen, W. Yu, R. Chen, and L. Lin, "Knowledge-embedded routing network for scene graph generation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019*, pp. 6163–6171, Long beach california, June 2019.
- [18] Z. X. Wang, T. S. Chen, S. J. Jimmy, W. Yu, H. Cheng, and L. Lin, "Deep reasoning with knowledge graph for social relationship understanding," 2018, <https://arxiv.org/abs/1807.00504v1>.
- [19] Q. Wu, P. Wang, C. Shen, D. Anthony, and A. van den Hengel, "Ask me anything: free-form visual question answering based on knowledge from external sources," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4622–4630, Las Vegas, NV, USA, June 2016.
- [20] Z. Wang, "Development status and trend outlook of edge computing," *Automation Panorama*, vol. 38, no. 2, pp. 22–29, 2021.
- [21] C. C. Juan, Z. Sherali, and G. Juan Antonio, "Internet of vehicles: architecture, protocols, and security," *IEEE Internet of Things Journal*, vol. 5, pp. 3701–3709, 2017.
- [22] W. S. Shi, X. Z. Zhang, Y. F. Wang, and Q. Zhang, "Edge computing: state-of-the-art and future directions," *Journal of Computer Research and Development*, vol. 56, no. 1, pp. 73–93, 2019.
- [23] Z. Liao, Y. Ma, J. Huang, J. Wang, and J. Wang, "HOTSPOT: a UAV-assisted dynamic mobility-aware offloading for mobile-edge computing in 3-D space," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10940–10952, 2021.
- [24] J. Zhang, J. Cui, H. Zhong, Z. Chen, and L. Liu, "PA-CRT: Chinese remainder theorem based conditional privacy-preserving authentication scheme in vehicular ad-hoc networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 722–735, 2021.
- [25] R. Roman, J. Lopez, and M. Mambo, "Mobile Edge Computing, Fog et al.: a Survey and Analysis of Security Threats and Challenges," *Future Generation Computer Systems*, vol. 78, pp. 680–698, 2018.
- [26] J. Zhou, X. L. Dong, and Z. F. Cao, "Research advances on privacy preserving in recommender systems," *Journal of Computer Research and Development*, vol. 56, no. 10, pp. 2033–2048, 2019.
- [27] Y. Z. Wu, Q. Liu, R. R. Chen, C. Li, and Z. Peng, "A group recommendation system of network document resource based on knowledge graph and LSTM in edge computing," *Security and Communication Networks*, vol. 2020, Article ID 8843803, , 2020.
- [28] D. Liu and Y. Luo, "A consensus reaching process based on the concordance correlation measure of intuitionistic fuzzy sets in multi-criteria decision making," *Journal of Intelligent & Fuzzy Systems*, pp. 1–16, 2021.
- [29] C. Cadena, L. Carlone, H. Carrillo et al., "Past, present, and future of simultaneous localization and mapping: toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [30] Y. Zhao, G. L. Liu, G. H. Tian et al., "A survey of visual SLAM based on deep learning," *Robot*, vol. 39, no. 6, pp. 889–896, 2017.
- [31] S. P. Li and T. Zhang, "A survey of deep learning application in visual SLAM," *Aerospace Control and Application*, vol. 45, no. 2, pp. 1–10, 2019.
- [32] R. J. Liu, X. S. Wang, and C. Zhang, "A survey on visual SLAM based on deep learning," *Journal of System Simulation*, vol. 32, no. 7, pp. 1244–1256, 2020.
- [33] H. Wang, W. Q. Zhu, Y. Z. Wu, P. J. He, and L. J. Wan, "Named entity recognition based on equipment and fault field of CNC machine tools," *Journal of Engineering Science*, vol. 42, no. 4, pp. 476–482, 2020.