applied sciences

MDPI

Article

A Kitchen Standard Dress Detection Method Based on the YOLOv5s Embedded Model

Ziyun Zhou [1], Chengjiang Zhou [2,*], Anning Pan [3,4,*], Fuqing Zhang [2], Chaoqun Dong [2], Xuedong Liu [2], Xiangshuai Zhai [4] and Haitao Wang [2]

[1] Information Center of Yunnan Administration for Market Regulation, Kunming 650228, China
[2] School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
[3] School of Big Data, Baoshan University, Baoshan 678000, China
[4] School of Physics and Electronic Information, Yunnan Normal University, Kunming 650500, China
* Correspondence: chengjiangzhou@foxmail.com (C.Z.); paninglw@163.com (A.P.)

Abstract: In order to quickly and accurately detect whether a chef is wearing a hat and mask, a kitchen standard dress detection method based on the YOLOv5s embedded model is proposed. Firstly, a complete kitchen scene dataset was constructed, and the introduction of images for the wearing of masks and hats allows for the low reliability problem caused by a single detection object to be effectively avoided. Secondly, the embedded detection system based on Jetson Xavier NX was introduced into kitchen standard dress detection for the first time, which accurately realizes real-time detection and early warning of non-standard dress. Among them, the combination of YOLOv5 and DeepStream SDK effectively improved the accuracy and effectiveness of standard dress detection in the complex kitchen background. Multiple sets of experiments show that the detection system based on YOLOv5s has the highest average accuracy of 0.857 and the fastest speed of 31.42 FPS. Therefore, the proposed detection method provided strong technical support for kitchen hygiene and food safety.

Keywords: standard dress; YOLOv5s embedded model; embedded detection system; food safety

Citation: Zhou, Z.; Zhou, C.; Pan, A.; Zhang, F.; Dong, C.; Liu, X.; Zhai, X.; Wang, H. A Kitchen Standard Dress Detection Method Based on the YOLOv5s Embedded Model. *Appl. Sci.* **2023**, *13*, 2213. https://doi.org/10.3390/app13042213

Academic Editor: Hui Yuan

Received: 25 November 2022
Revised: 14 January 2023
Accepted: 15 January 2023
Published: 9 February 2023

1. Introduction

To solve the problem of food safety [1], the CFDA has launched a "transparent kitchen and stoves" campaign, showing the details of cooking through video displays, open kitchens, and other ways [2]. The latest Food Safety Law [3] also stipulates that the relevant personnel should wear clean work clothes and hats when cooking [4]. Food safety standards propose that the wearing standards for chefs must be standardized strictly [5]. Through the standardization of chefs' wearing standards, the safety levels of food will be improved, and the risk of disease transmission and food poisoning will be reduced. Chefs should wear chef hats, aprons, and masks, so the kitchen standard dress includes the chef's hat, apron, and mask. Assuming that the chef does not wear a chef's hat and apron, the dirty things on the body of the chef will pollute the food. Assuming the chef does not wear a mask, the droplets from the chef's breath or coughing will fall onto the food. If the chef is suffering from a disease, the risk of virus transmission will be increased greatly [6]. Therefore, kitchen standard dress detection is particularly important.

At present, there are only a few studies on kitchen wear detection, and these studies are based on target detection. Target detection methods can be divided into one-phase and two-stage methods. Two-stage networks include CenterNet [7], R-CNN, and Faster R-CNN [8], etc. The one-stage detection network includes a single shot detector (SSD) [9], YOLO, etc. Guo [10] proposed a behavior monitoring model of kitchen staff based on YOLOv5l and DeepSort [11], which can detect whether chefs wear hats and masks, but it only uses 2000 pictures to train the model and has low efficiency and accuracy. Tao [12] proposed a system with a deep learning method to automatically identify the violation of

kitchen staff, including not wearing a chef's hat and smoking. Ramadan [13] used long short-term memory and the hidden Markov model to detect cooking action in the cooking process, and the recognition rate was 81% and the speed was 35 frames. Sheng [1] proposed a scheme for cooking assistants' overalls based on the Hi3559A embedded processor, in which the speed of overalls detection is improved by YOLOv3 network optimization and parallel processing technology, and the recognition speed is 28 frames. Staden [14] detected the hands of operators in kitchens through YOLOv3 [15] based on MobileNet-lite and VGG16 but the detection speed and accuracy under a complex background are not satisfactory. Lu [16] improved the Faster R-CNN [17] through the deeper region proposal network (D-RPN), and reconstructed the U-network through the feature enhancement module; the text recognition accuracy on the kitchen electrical control panel reached 89.84%. Traditional detection methods and a small number of one-stage and two-stage target detection algorithms have been applied in kitchen wear detection, but these methods are not only outdated but they also have some inherent defects. The two-stage object detection algorithm generates a large number of proposal boxes, which results in the relatively slow speed of two-stage algorithms such as R-CNN, Fast R-CNN, and Faster R-CNN. Single stage detection methods such as SSD, YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6, and YOLOv7 have simple structures, fast inference speeds, and convenient deployment methods, but their accuracy is slightly lower than that of two-stage-based methods [6]. Wang proposed a fast detection method of the cannibalistic behavior of juvenile fish based on YOLOv5 [18], in which the backbone network is replaced by the multi head attention mechanism. Li [19] proposed a workers' personal protective equipment detection method based on YOLOv5, which can improve the efficiency of safety management. In addition, FCOS and transformer-based methods have recently been used for object detection. Tan adopts an improved FCOS to recognize the numbers in the industrial instruments; the balanced feature pyramid is combined with FPN to further integrate the feature information [20]. Deshmukh proposed a Swin-Transformer-based vehicle detection framework in an undisciplined traffic environment [21]. For research on kitchen dress codes, the current study only focuses on hat wearing and smoking behavior, and the number of datasets is small, which leads to low detection accuracy and reliability. In addition, the embedded deployment of detection models is necessary, but this is still a research gap. For detection performance, the kitchen wear detection methods based on Faster R-CNN, YOLOv3, and VGG16 are old and simple, while the latest methods based on FCOS and Swin Transformer have complex network structures, which makes the selection and improvement of the methods more difficult. In our experiments, YOLOv5s is not only fast but also highly accurate, so it is considered to be used for kitchen wear detection. However, it is still difficult for the existing methods to meet the accuracy and real-time requirements. The following problems still exist in the research of kitchen wear detection.

(1) Most of the existing kitchen wear detection methods only detect one kind of wearable clothing, which cannot meet the requirements of kitchen wear detection.

(2) Chef hats, aprons, and masks all come in different styles, different shapes, and different colors, and it is especially difficult to identify chef hats, aprons, and masks at the same time under different backgrounds.

(3) The detection process currently in use consists of the following steps. Firstly, image or video data are captured via a camera and the video signal is transmitted to the server over the network. Then, illegal wearing is detected by a detection model deployed on the server, and the detection results and alarm signals are transmitted to the client. However, this detection process is time-consuming, and the detection delay, network delay, and alarm delay will reduce the effectiveness and reliability of kitchen wear condition detection. Therefore, how to reduce the detection delay, network delay, and alarm delay is an urgent problem to be solved.

(4) The display, lighting, and background of items in different kitchens are different, and traditional methods, YOLOv3, and Faster R-CNN cannot detect the wearing of kitchen clothes under complex backgrounds and complex lighting.

In order to improve the accuracy, reliability, and real-time standard detection of the wearing of hats, masks, and aprons under complex backgrounds and multiple operating conditions, a set of rich kitchen wearing detection datasets and an embedded detection model are our research scope, which will provide AI technical support for kitchen hygiene and food safety. The main contributions and innovations of the paper are summarized as follows.

(1) A Kitchen Standard Dress Detection Method Based on the YOLOv5s Embedded Model is proposed, which can detect whether a chef is wearing a chef's cap and mask accurately and quickly, not just a single mask or hat detection. Several improvements to YOLOv5s have improved the accuracy of wearing detection under complex backgrounds and lighting, and the YOLOv5s embedded model further reduces the detection delay.

(2) An embedded detection system based on Jetson Xavier NX and YOLOv5s is proposed, which minimizes network transmission delay and early warning delay. The embedded detection system is not only cheap, but can also be deployed directly in the kitchen to detect illegal wearing behavior and send an alarm on the first occassion.

(3) A complete dataset of kitchen wearing case images under multiple kitchen scenes is proposed. The dataset includes 7558 images, different kitchen scenes, different chef hats, and different mask styles, which means that our detection method can achieve different types of dress detection, rather than a single hat or mask detection.

The rest of this paper is arranged as follows. The standard dress detection system is described in Section 2. Experimental results and analysis are included in Section 3. Finally, we present the experimental results and related discussions in Section 4.

2. The Standard Dress Detection System

2.1. Detection Model Training Stage

The standard dress detection system based on the Jetson Xavier NX can be implemented in two stages, including the detection model training stage and the Jetson Xavier NX model deployment and application stage. In the detection model training stage, the smallest YOLOv5s model with good performance is trained to detect whether a chef is wearing both a hat and a mask. The purpose of this phase is to train and generate the "pt" model with the best detection performance, which does not involve the connection and operation of the Jetson Xavier NX.

To ensure the reliability of kitchen standard dress detection, a target detection algorithm with high accuracy and reliability is necessary. YOLOv5 is a relatively new object detection algorithm with good performance. According to the different network widths and depths, four different versions have been released, including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. To ensure the real-time and accurate detection of standard dress, we used the smallest model, YOLOv5s, as shown in Figure 1. The YOLOv5s model consists of four parts, input, backbone, neck, and prediction [22].

(1) Input: The input end is composed of Mosaic data augmentation, adaptive anchor frame calculation, adaptive image scaling, and other small modules. Mosaic data enhancement includes random scaling, random clipping, and random arrangement processing, the adaptive anchor frame calculation can adaptively calculate the best anchor frame in different training sets, and adaptive image scaling can improve the inference speed.

(2) Backbone: The backbone is mainly composed of a focus structure and CSP structure. The key to the focus structure is the slicing operation, and the 4*4*3 image becomes a 2*2*12 feature map after slicing. In fact, there are two CSP structures designed in Yolov5, the backbone network of which uses a CSP1_X structure.

(3) Neck: Both Yolov5's neck and Yolov4's neck use FPN + PAN architecture, and other parts of the network have been adjusted. In the neck structure of Yolov4, ordinary convolution operations are used. However, Yolov5's neck structure borrows from CSPnet's CSP2 structure, which strengthens the ability of network feature fusion. YOLOv5s uses CSP1_X and CSP2_X to improve the speed of the network while ensuring the precision. The neck of YOLOv5s reduced the number of model parameters by the CSP2_X module and upsampled feature maps of 80*80*512 in size.

(4) Prediction: The prediction stage mainly includes two parts, the bounding box loss function and the NMS non-maximum suppression. GIOU loss is used as the loss function of the bounding box in YOLOv5. In the post-processing of the YOLOv5 target detection model, NMS non-maximum suppression is used to screen effective target frames.

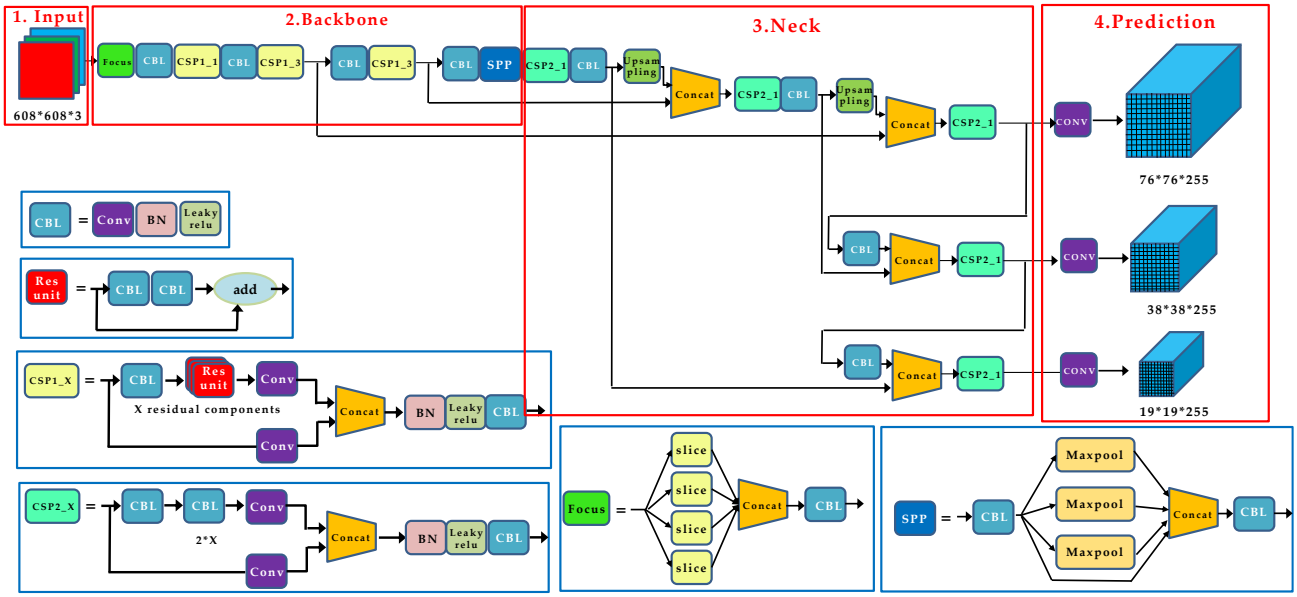

Figure 1. The YOLOv5s network structure.

In order to ensure the reliability and accuracy of standardized dress detection in actual kitchens, we need to pay attention to two particularly critical issues in the training stage, including the universality of the dataset and the parameter consistency of model training. For the training dataset problem, we collected a large number of videos and images of real kitchen scenes, including various dress situations that can be encountered in kitchen cooking, which can meet the needs of model training and ensure the accuracy of dress detection in real scenarios. For the problem of parameter consistency, we trained various detection models through the same super parameters, the same pre-training model, and the same hardware devices, which ensures the reliability of the model and the fairness of comparative experiments. After training, the "pt" model file that can be used for the deployment of Jetson Xavier NX embedded devices is available.

2.2. Model Deployment and Application Stage Based on Jetson Xavier NX

In a conventional detection system, images or videos are captured via a camera and the video signal is transmitted to the server over the network. Then, non-standard dress is detected by a detection model deployed on the server, and the detection results and alarm signals are transmitted to the client. The process is time-consuming, and the network delay and alarm delay will reduce the reliability of standard dress detection. As a result, the trained YOLOv5 model was deployed to the Jetson Xavier NX embedded device, and the embedded device was deployed in a wide variety of kitchens. With the support of the powerful computing power of Jetson Xavier NX, the server deployment and network transmission are omitted, and real-time detection and early warning of non-standard clothing become possible.

Jetson Xavier NX can be used as an edge server or as a terminal device, so many computing tasks can be implemented separately based on it [23]. Jetson Xavier NX is small in size, portable, and powerful in computing performance, which makes it suitable for embedding into UAVs [24], portable medical devices [25], small robots [26], and other embedded systems of the Internet of Things [27]. As shown in Table 1, we provide some parameters of the Jetson Xavier NX device, which indicate that the embedded device is low in energy consumption,

high in computing efficiency, and good in model portability. Jetson Xavier NX supports the deep learning model, which meets the performance requirements of edge computing terminals for object detection [28]. In addition, there is a WIFI module on the back of the Jetson Xavier NX, which can connect to the host system through wireless networking.

Table 1. Parameters of Jetson Xavier NX.

Items	Parameters
GPU	384-core NVIDIA Volta^{TM} architecture with 48 Tensor cores
CPU	six-core NVIDIA Carmel ARMv8.2 64-bit CPU
Memory	16 GB 128-bit LPDDR4x@1600MHz
Data storage	16GB eMMC5.1
CSI	CSI support for up to six cameras (14 channels) MIPI CSI-2 D-PHY 1.2
PCIE	1 × 1(PCIE3.0) + 1 × 4(PCIE4.0), 144GT/s
Video encoding/ decoding	2K × 4K60 Hz encoding (HEVC); 2K4K60 Hz decoding
Size	69.6 mm × 45 mm

The process of deploying a detection model to Jetson Xavier NX can be divided into two key phases, as shown in Figure 2, including Jetson Xavier NX hardware deployment and DeepStream-Yolo Software Deployment. During the hardware deployment phase, the Jetson Xavier NX embedded device is connected to the server via a USB interface, which guarantees the fast transfer of packages, datasets, and configuration files. In addition, Jetson Xavier NX requires a separate power supply, and Jetson Xavier NX is connected to the server network through the LAN port, which ensures network connectivity for the installation and configuration of the environment. However, it is not necessary for the camera to be connected to the Jetson Xavier NX device during model deployment.

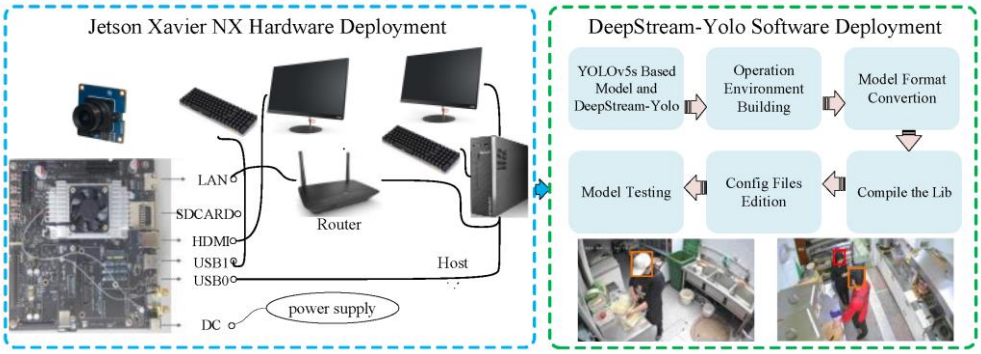

Figure 2. The process of deploying a detection model to Jetson Xavier NX.

During the DeepStream-Yolo software deployment, the deployment of YOLOv5s on the Jetson Xavier NX device is mainly based on the NVIDIA DeepStream SDK. Firstly, we need to set up the necessary running environment for the DeepStream YOLO and YOLOv5s models and prepare some video streams and images for testing. Subsequently, import a trained "pt" model, convert it to obtain the cfg and wts files, and then compile the lib and edit the configuration files according to the situation. Combined with different model accuracies in the training stage and the frame rates obtained from running on DeepStream, the final accuracy used is FP32. The average frame rates of different models under different accuracies are shown in Table 2. Finally, we tested the model and applied it to actual kitchen wear detection. It is worth noting that the training process is particularly important, so we used a high-performance host to train the detection model that will be embedded in the Jetson Xavier NX device.

Table 2. Average frame rate of different models on DeepStream with FP32, FP16, and INT8 accuracy.

Model	FP32(FPS)	FP16(FPS)	INT8(FPS)
YOLOv5s	31.56	60	61.05
YOLOv5x	3.94	12.1	12.43
YOLOv7	7.46	22.42	23.65
YOLOv7-tiny	39.12	71.2	72.08
YOLOv7x	4.06	12.8	13.2
YOLOv7e6	6.05	18.64	19.7

After the trained YOLOv5.pt model is successfully deployed to Jetson Xavier NX, we can use the current Jetson Xavier NX embedded device as a complete kitchen standard dress detection system, that is, it can be deployed to any kitchen without the host. As shown in Figure 3, a kitchen normalized dress detection system based on the YOLOv5s embedded model is deployed in a relatively high corner of the kitchen, which can detect whether a chef is wearing a mask and hat in real time without a network.

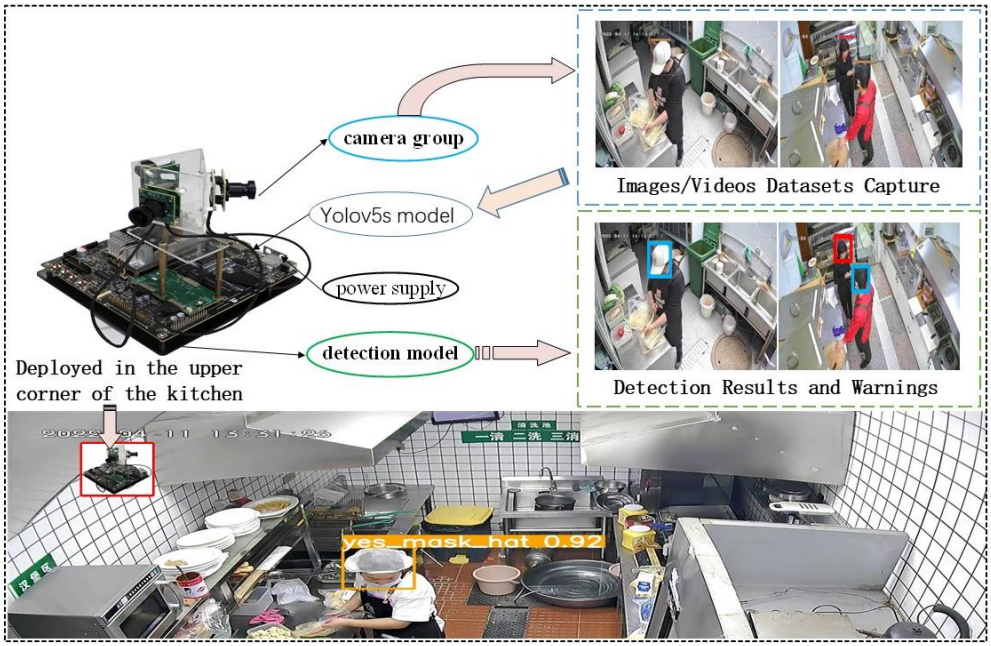

Figure 3. The process of deploying a standard dress detection system to a kitchen.

According to the processing process, the proposed detection method for kitchen standard dress detection includes three stages, as shown in Figure 3.

(1) During the data acquisition phase, we deployed the Jetson Xavier NX embedded device to the kitchen and capture kitchen video streams or images in real time via a monocular camera connected to the Jetson Xavier NX device.

(2) In the image pre-processing phase, we process the video streams and images via Jetson Xavier NX with the deployment of the YOLOv5s training model. The powerful computing power of Jetson Xavier NX ensures the accuracy and real-time performance of standardized dress detection in the kitchen, and this process does not require the support of the host.

(3) In the stage of detection result display and non-standard dress warning, a high-definition display and voice warning device can be connected to the USB terminal or the wireless network of the Jetson Xavier NX, which can monitor the chef's non-standard dress in real-time in the form of videos or voice.

3. Experiments and Discussion

3.1. Experiment Setup and Data Preparation

To verify the effectiveness of proposed kitchen standard dress detection method, we compared and analyzed the performance of detection models through a large number of experiments. Our experiments are based on the PyTorch framework, which is deployed on the Ubuntu operating system. In the process of training models, we used a computer with an NVIDIA GeForce RTX 3080 Ti GPU, whose memory is 12 GB. In the end, the edge device was used to measure the speed and the performance while detecting the test sets. As shown in Table 3, we implemented the training and testing for the models in the following experimental environment.

Table 3. Experimental environment configuration.

Parameter	Configuration
CPU	Intel Core i7-10700F, CPU 2.90 Hz, RAM 32 GB
GPU	Nvidia GeForce GTX 2080Ti (24 G)
Accelerated Environment	CUDA 11.1, cuDNN8.0.5
Visual Studio System	Pytorch1.7.1, Python 3.6
Operating System	Ubuntu 18.04

In order to obtain sufficient and realistic datasets, we collected 7558 kitchen images through cameras installed in various restaurants in Chengjiang City, Yuxi City, which provides powerful data support for kitchen standard dress detection. The 7558 images contained four categories, including no mask and no hat (labeled no_mask_hat), mask and hat (labeled yes_mask_hat), no mask and hat (labeled no_mask_yes_hat), and mask and no hat (labeled yes_mask_no_hat).

As shown in Figure 4, these kitchen images were collected under different illuminations, distances, and backgrounds, so the size and angle of the chef who appears in the kitchen are different. The kitchen background is particularly complex in many images, which makes it particularly difficult to detect the chef's wearing behavior from the background. We divided the 7558 images into training and validation sets with sizes of 6047 and 1511, respectively, close to 4:1. A test set was also made up of 200 images captured in the real situation with about the four classes.

Figure 4. Kitchen standard dress detection images in complex environments.

To verify the validity of the proposed methods, we evaluated the detection performance of various methods through precision (P), recall (R), F1 score (F1), mean average precision (mAP), and speed. Moreover, the detailed definitions of precision, recall, F1 score, and mean average precision can be found in the literature [21]. Among them, F1 represents the harmonic mean of precision and recall. The closer the parameter is to one, the better

the detection performance of the model. Average precision (AP) represents the area under the precision recall curve. mAP refers to the average recognition accuracy of all of the categories and speed represents the sum of the pre-processing time, inference time, and non-maximum suppression time in Jetson Xavier NX.

3.2. Performance Comparison with Existing Methods

The reason why we put forward the detection system is to improve the accuracy of identifying whether chefs are wearing masks and hats and to make sure the cooking process is standardized and the food hygiene is safe. In the experiment, we utilized a series of YOLOv5 and YOLOv7 object detectors to find one with good speed and remarkable detection performance. We trained six state-of-the-art object detection networks under the same parameters, the same kitchen standard dress detection dataset, and a fair experimental environment, including YOLOv5-s, YOLOv5-x, YOLOv7, YOLOv7-tiny, YOLOv7x, and YOLOv7-e6. All of the models were trained based on their pre-training weights; we used the same datasets we constructed before for training and validation. The detailed training parameter settings are shown in Table 4.

Table 4. Training parameter settings of YOLOv5 and YOLOv7.

Epoch	Batch Size	IoU Threshold	Initial Learning Rate	Momentum	Input Image
500	8	0.2	0.01	0.937	640*640

It can be seen from Figure 5 that when the training set was trained to 300 epochs, the training accuracy curves of all of the models tended to converge. Among them, YOLOv5s (red line) not only has good convergence performance for accuracy curves, but it also has higher training accuracy. At the end of training, YOLOv7x (sky blue line) achieved the highest accuracy. As shown in Figure 6, except for YOLOv5x, YOLOv5s (red line) has maintained the highest recall rate, indicating that it has better convergence speed and accuracy. However, the recall of YOLOv7x, which has the highest training accuracy, is not high, and the network parameters and time cost on YOLOv7x are huge. YOLOv5s keeps a high recall rate before the 400 epochs, indicating that its convergence speed and accuracy are better.

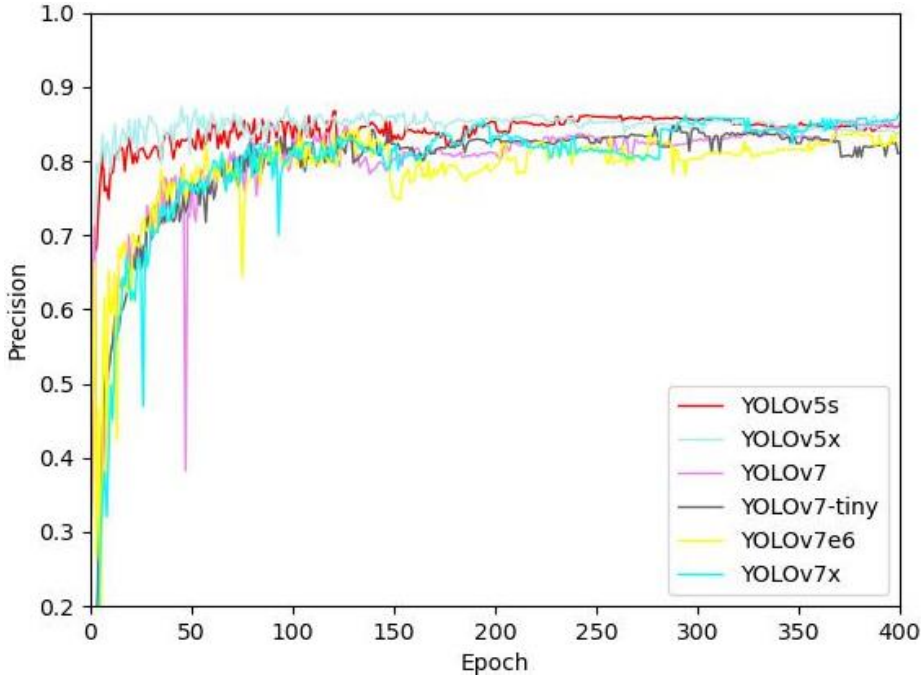

Figure 5. Comparison of training precision.

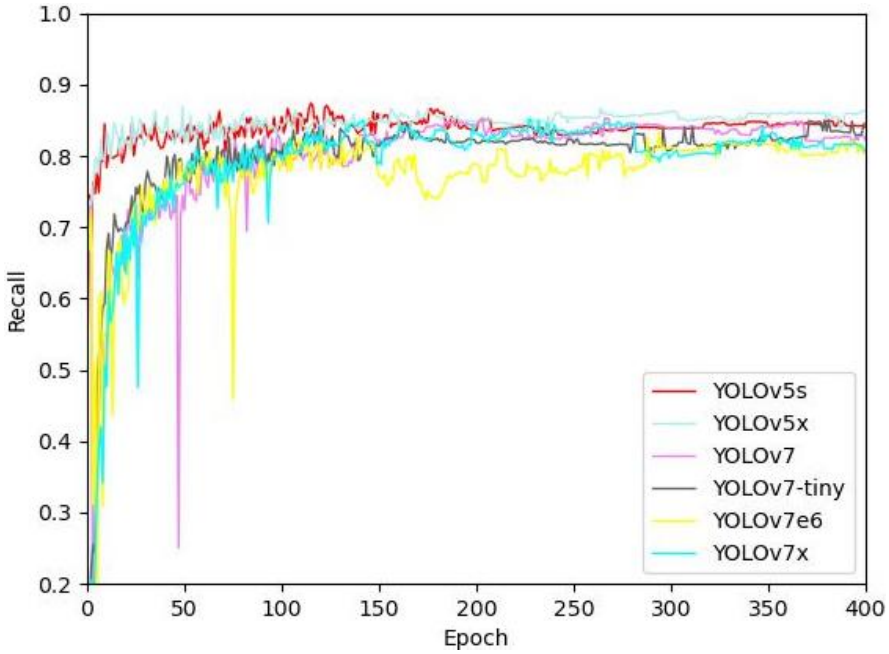

Figure 6. Comparison of training recall.

In order to make the Jetson Xavier NX embedded device implement a good detection effect on the clothing wear of kitchen staff, we analyzed and compared it with the evaluation criteria of six models on the basis of the above experimental environment and then selected the model most suitable for the Jetson Xavier NX embedded device. The precision (P) and recall (R) of the four conditions for chefs' clothing wear, no mask and no hat (labeled no_mask_hat), mask and hat (labeled yes_mask_hat), no mask and hat (labeled no_mask_yes_hat), and mask and no hat (labeled yes_mask_no_hat), are shown in Table 5. Compared with no_mask_yes_hat, there are more images and cases for the other three classes. In the three cases, their accurate detection is easy while the label no_mask_yes_hat is a bit hard due to the lack of images to train. Moreover, YOLOv5s obtained the best precision in three categories: no_mask_hat, yes_mask_hat, and yes_mask_no_hat.

Table 5. The precision (P) and recall (R) of the four conditions for chefs' dressing.

Class	Labels	YOLOv5s		YOLOv5x		YOLOv7		YOLOv7-tiny		YOLOv7x		YOLOv7e6	
		P	R	P	R	P	R	P	R	P	R	P	R
no_mask_hat	728	0.867	0.852	0.858	0.878	0.855	0.856	0.828	0.852	0.86	0.854	0.853	0.832
yes_mask_hat	1064	0.919	0.886	0.907	0.907	0.899	0.902	0.897	0.898	0.891	0.91	0.917	0.875
no_mask_yes_hat	346	0.772	0.775	0.779	0.795	0.753	0.803	0.738	0.783	0.771	0.775	0.751	0.876
yes_mask_no_hat	790	0.868	0.784	0.86	0.824	0.845	0.81	0.833	0.805	0.848	0.797	0.842	0.773

In order to further analyze the detection effect of various kitchen dressing statuses, we also provided a YOLOv5s-based classification confusion matrix, as shown in Figure 7. The "yes_mask_hat" state has the highest detection accuracy of 0.91, but this state is easily recognized as a background state, that is, a small number of cases cannot be recognized. The detection accuracy of the "yes_mask_no_hat", "no_mask_hat", and "no_mask_yes_hat" states are 0.84, 0.83, and 0.79, respectively. These three states are also easy to recognized as background states, but there is no confusion between various kitchen dressing states under the condition of the YOLOv5s detection model. It is worth noting that there are 790, 728, and 346 samples in the "yes_mask_no_hat", "no_mask_hat", and "no_mask_yes_hat" states, respectively, which indicates that the accuracy of the detection model depends largely on the number of samples. Therefore, it is very meaningful for us to collect 7558 images through multiple cameras.

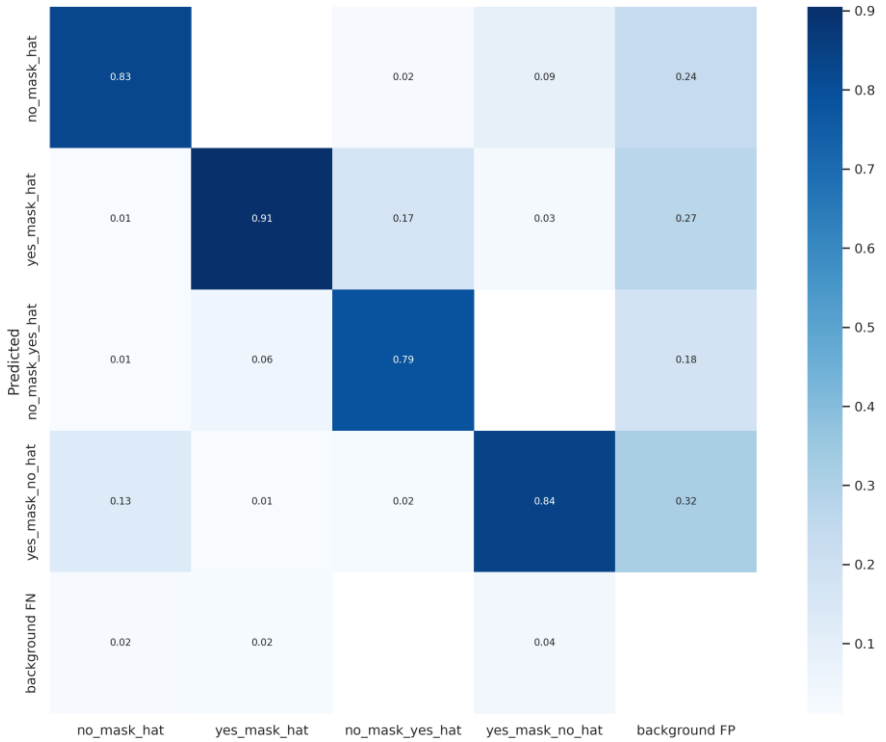

Figure 7. The confusion matrix obtained from various kitchen dress states.

In addition, the detection results of the above six models in the testing datasets are shown in Table 6 and Figure 8. It can be seen that YOLOv5s has the highest precision and the second highest speed of 315 ms per image, and YOLOv7 has the best recall, mAP, and F1 score. Thus, the images detected by YOLOv5s have better precision, especially for the class no_mask_hat which is represented with red-colored boxes. As for the other classes, YOLOv5x is the only competitor, but it is quite a lot slower than YOLOv5s. Detailed images are shown in Figure 8.

Table 6. Performance of different detection models.

Model	P	R	mAP@.5	mAP@.5:.95	F_1	Speed in Jetson (ms)
YOLOv5s	0.857	0.824	0.862	0.618	0.840	315
YOLOv5x	0.851	0.851	0.856	0.618	0.851	2694
YOLOv7	0.855	0.856	0.891	0.629	0.855	1570
YOLOv7-tiny	0.824	0.835	0.879	0.613	0.830	212
YOLOv7x	0.842	0.834	0.886	0.632	0.838	2467
YOLOv7e6	0.841	0.817	0.873	0.623	0.829	2063

3.3. Results in the Kitchen

In real life, it is necessary to quickly and accurately judge and warn of non-standard wearing behaviors in kitchens. Therefore, it is crucial to embed a model with high detection accuracy and speed into the Jetson Xavier NX portable device to achieve a good detection effect in the actual scene. Through the above experimental comparison and analysis, it can be concluded that YOLOv5s can achieve a high detection speed of 33 FPS while maintaining good detect accuracy, which can meet the requirements of the model in this paper. Therefore, we choose to embed the YOLOv5s model into Jetson to detect the standard behavior of kitchen staff in an actual working environment.

To demonstrate the performance of our trained YOLOv5s model, we evaluated it with the testing datasets constructed by images taken from several surveillance cameras. We transferred the test set to our edge device to measure the speed of the entire system. The testing performance is shown in Figure 9. As we can see in Figure 9, our method has good

performance in both each of the classes and overall in different situations. The results show that our final system can detect at a speed of 315 ms per image and 31.42 FPS for videos in our edge device and the confidence has reached the practical application requirements. However, there is still some low-quality and incorrect detection due to the unexpected behaviors, such as wearing plastic head covers instead of hats, wearing masks lacking in standardization, and so on. In the future, we will improve the detection network structure and modify the algorithm logic of violation detection to make the performance better.

Figure 8. The detection performance of different models. It should be noted that the colors orange, green, red, and copper represent the labels yes_mask_hat, yes_mask_no_hat, no_mask_hat, and no_mask_yes_hat, respectively.

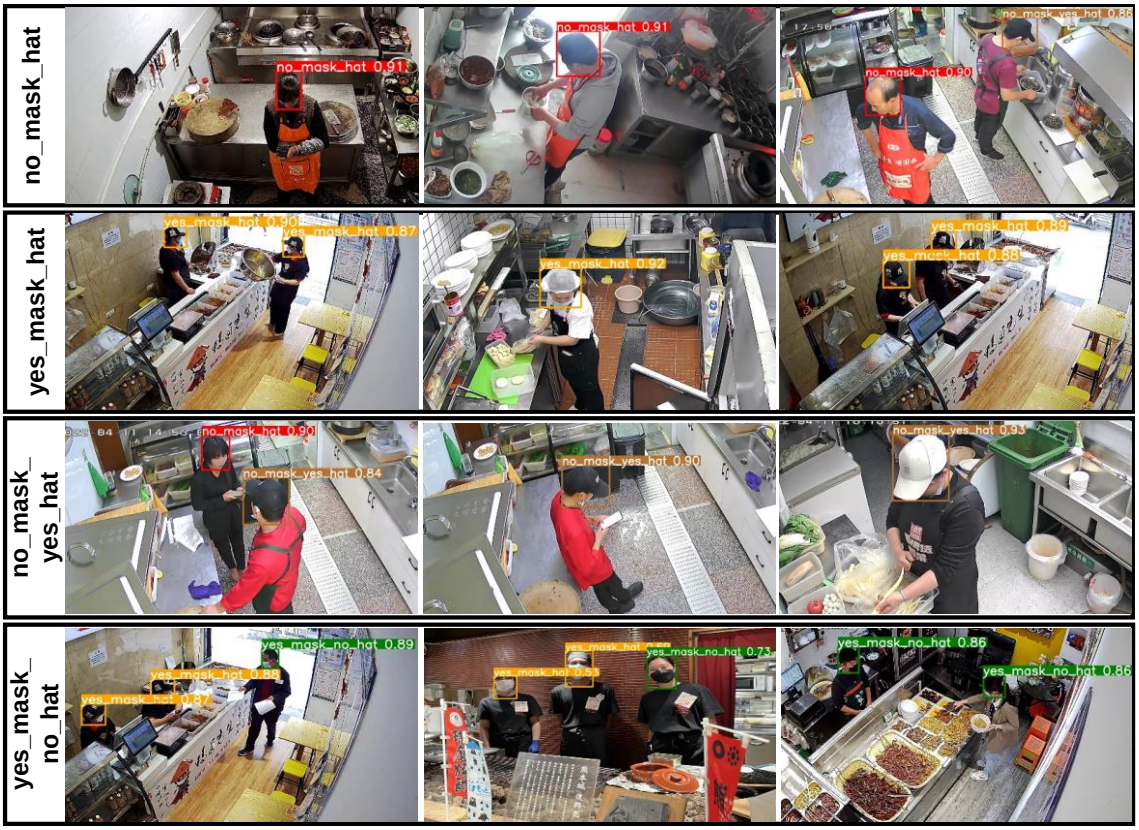

Figure 9. The detection performance of the YOLOv5s embedded model. It should be noted that the colors orange, green, red, and copper represent the labels yes_mask_hat, yes_mask_no_hat, no_mask_hat, and no_mask_yes_hat, respectively.

4. Conclusions

In order to ensure that chefs wear masks and hats correctly, this paper proposes a kitchen staff dress monitoring system so that food can be kept clean. The system, based on the YOLOv5s model, automatically detects violations in the back of the kitchen and saves the violation clips. We have been permitted to utilize the surveillance cameras of several stores and restaurants to collect datasets and fortunately, our trained system performs well in these numerous situations. The experimental results show that the kitchen staff dress detection system we proposed can detect violations in the practical back of a kitchen with good accuracy and robustness.

Author Contributions: Conceptualization, Z.Z. and C.Z.; methodology, Z.Z. and C.Z.; software, A.P.; validation, A.P., F.Z. and C.D.; formal analysis, X.L.; investigation, X.Z.; resources, Z.Z., C.Z. and A.P.; data curation, H.W.; writing—original draft preparation, F.Z.; writing—review and ed-iting, Z.Z. and A.P.; visualization, C.Z.; supervision, Z.Z. and C.Z.; project administration, Z.Z. and C.Z.; funding acquisition, Z.Z. and C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Yunnan Provincial Market Supervision Bureau IOT perception platform construction project, the PhD research startup foundation of Yunnan Normal University (No. 0100020502 0503131).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments: The author sincerely thanks the team for their guidance. The author sincerely expresses thanks to the reviewers for taking the time to review the paper in their busy schedule.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sheng, Q.; Sheng, H.; Gao, P.; Li, Z.; Yin, H. Real-Time Detection of Cook Assistant Overalls Based on Embedded Reasoning. *Sensors* **2021**, *21*, 8069. [CrossRef] [PubMed]
2. Ventrella, J.; MacCarty, N. Monitoring impacts of clean cookstoves and fuels with the Fuel Use Electronic Logger (FUEL): Results of pilot testing. *Energy Sustain. Dev.* **2019**, *52*, 82–95.
3. Geng, S.; Liu, X.; Beachy, R. New Food Safety Law of China and the special issue on food safety in China. *J. Integr. Agric.* **2015**, *14*, 2136–2141. [CrossRef]
4. Mihalache, O.A.; Møretrø, T.; Borda, D.; Dumitraşcu, L.; Neagu, C.; Nguyen-The, C.; Maître, I.; Didier, P.; Teixeira, P.; Lopes Junqueira, L.O.; et al. Kitchen layouts and consumers' food hygiene practices: Ergonomics versus safety. *Food Control* **2022**, *131*, 108433. [PubMed]
5. Chang, H.; Capuozzo, B.; Okumus, B.; Cho, M. Why cleaning the invisible in restaurants is important during COVID-19: A case study of indoor air quality of an open-kitchen restaurant. *Int. J. Hosp. Manag.* **2021**, *94*, 102854. [CrossRef]
6. Jewitt, S.; Smallman-Raynor, M.; K C, B.; Robinson, B.; Adhikari, P.; Evans, C.; Karmacharya, B.M.; Bolton, C.E.; Hall, I.P. Domesticating cleaner cookstoves for improved respiratory health: Using approaches from the sanitation sector to explore the adoption and sustained use of improved cooking technologies in Nepal. *Soc. Sci. Med.* **2022**, *308*, 115201.
7. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6569–6578.
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *39*, 1137–1149.
9. Ahmed, M.D.F.; Mohanta, J.C.; Sanyal, A. Inspection and identification of transmission line insulator breakdown based on deep learning using aerial images. *Electr. Power Syst. Res.* **2022**, *211*, 108199.
10. Guo, X.; Zuo, M.; Yan, W.; Zhang, Q.; Xie, S.; Zhong, I. Behavior monitoring model of kitchen staff based on YOLOv5l and DeepSort techniques. In Proceedings of the MATEC Web of Conferences, Xiamen, China, 29–30 December 2021; pp. 1–7.
11. Lin, Y.; Chen, T.; Liu, S.; Cai, Y.; Shi, H.; Zheng, D.; Lan, Y.; Yue, X.; Zhang, L. Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning. *Comput. Electron. Agric.* **2022**, *197*, 106938. [CrossRef]
12. Tao, L.; Ruixia, W.; Biao, C.; Jianlin, Z. Implementation of kitchen food safety regulations detection system based on deep learning. In Proceedings of the 2021 6th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Oita, Japan, 25–27 November 2021; pp. 59–62.
13. Ramadan, M.; El-Jaroudi, A. Action detection and classification in kitchen activities videos using graph decoding. *Vis. Comput.* **2022**, 1–14. [CrossRef]
14. van Staden, J.; Brown, D. An Evaluation of YOLO-Based Algorithms for Hand Detection in the Kitchen. In Proceedings of the 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 5–6 August 2021; pp. 1–7.
15. Yan, J.; Wang, Z. YOLO V3 + VGG16-based automatic operations monitoring and analysis in a manufacturing workshop under Industry 4.0. *J. Manuf. Syst.* **2022**, *63*, 134–142. [CrossRef]
16. Lu, M.; Chen, L. Efficient object detection algorithm in kitchen appliance scene images based on deep learning. *Math. Probl. Eng.* **2020**, *2020*, 6641491. [CrossRef]
17. Jiang, Q.; Jia, M.; Bi, L.; Zhuang, Z.; Gao, K. Development of a core feature identification application based on the Faster R-CNN algorithm. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105200. [CrossRef]
18. Wang, H.; Zhang, S.; Zhao, S.; Lu, J.; Wang, Y.; Li, D.; Zhao, R. Fast detection of cannibalism behavior of juvenile fish based on deep learning. *Comput. Electron. Agric.* **2022**, *198*, 107033.
19. Li, J.; Zhao, X.; Zhou, G.; Zhang, M. Standardized use inspection of workers' personal protective equipment based on deep learning. *Saf. Sci.* **2022**, *150*, 105689. [CrossRef]
20. Tan, Y.; Yu, D.; Hu, Y. An application of an improved FCOS algorithm in detection and recognition of industrial instruments. *Procedia Comput. Sci.* **2021**, *183*, 237–244.
21. Deshmukh, P.; Satyanarayana, G.S.R.; Majhi, S.; Sahoo, U.K.; Das, S.K. Swin transformer based vehicle detection in undisciplined traffic environment. *Expert Syst. Appl.* **2023**, *213*, 118992. [CrossRef]
22. Ying, Z.; Lin, Z.; Wu, Z.; Liang, K.; Hu, X. A modified-YOLOv5s model for detection of wire braided hose defects. *Measurement* **2022**, *190*, 110683.
23. Li, Y.; Wang, J.; Wu, H.; Yu, Y.; Sun, H.; Zhang, H. Detection of powdery mildew on strawberry leaves based on DAC-YOLOv4 model. *Comput. Electron. Agric.* **2022**, *202*, 107418. [CrossRef]
24. Han, G.; He, M.; Zhao, F.; Xu, Z.; Zhang, M.; Qin, L. Insulator detection and damage identification based on improved lightweight YOLOv4 network. *Energy Rep.* **2021**, *7*, 187–197.

25. Zhang, P.; Wang, C.; Jiang, C.; Han, Z. Deep reinforcement learning assisted federated learning algorithm for data management of IIoT. *IEEE Trans. Ind. Inform.* **2021**, *17*, 8475–8484. [CrossRef]
26. Qiu, C.; Aujla, G.S.; Jiang, J.; Wen, W.; Zhang, P. Rendering Secure and Trustworthy Edge Intelligence in 5G-Enabled IIoT using Proof of Learning Consensus Protocol. *IEEE Trans. Ind. Inform.* **2022**, *19*, 9789427.
27. Zhang, P.; Jiang, C.; Pang, X.; Qian, Y. STEC-IoT: A security tactic by virtualizing edge computing on IoT. *IEEE Internet Things J.* **2020**, *8*, 2459–2467. [CrossRef]
28. Li, Q.; Zhao, F.; Xu, Z.; Li, K.; Wang, J.; Liu, H.; Qin, L.; Liu, K. Improved YOLOv4 algorithm for safety management of on-site power system work. *Energy Rep.* **2022**, *8*, 739–746. [CrossRef]