**International Journal of Digital Crime and Forensics**
Volume 16 • Issue 1

# Laboratory Dangerous Operation Behavior Detection System Based on Deep Learning Algorithm

Dawei Zhang, School of Information Engineering, Liao Dong University, China*

## ABSTRACT

Aiming at the problem that dangerous operation behaviors in the laboratory is difficult to identify by monitoring the video. An algorithm of dangerous operation behavior detection in multi-task laboratory based on improved YOLOv5 structure is proposed. Firstly, the algorithm enhances, adaptively scales, and adaptively anchors box computing on the input of YOLO network. Then convolution operation is carried out to strengthen the ability of network feature fusion. Finally, the GIoU_Loss function is used at the output to optimize the network parameters and accelerate the convergence of the model. The experimental results show that the algorithm performs well in real-time head localization, head segmentation, and population regression, with significant innovation and superiority. Compared with traditional methods, this algorithm has better accuracy and real-time performance and can more effectively achieve human operation behaviors detection in laboratory application environments.

## KEYWORDS

Behavior Detection, Deep Learning, Laboratory, YOLOv5

## INTRODUCTION

Human behavior is an important part of human life. There are many kinds of hazardous chemicals in university laboratories, which are scattered in storage locations and densely used (Dewi, 2021) and they are flammable, explosive, toxic, infectious and corrosive, and some high-risk chemicals even have highly toxic characteristics. If we can quickly identify dangerous behaviors by monitoring the video taken by the camera in the laboratory, we can find potential accident risks in time and deal with them immediately, which can effectively reduce or eliminate laboratory safety problems and ensure personal and property safety. Behavior detection and recognition has attracted more and more attention from relevant researchers because of its great application potential in monitoring system, video analysis and other fields (Ophoff et al., 2020; Li et al., 2021), and how to solve various problems in video behavior detection and recognition by using DL (Deep learning) technology is the hottest topic (Chen et al., 2021; Xu et al., 2020).

At present, the target detection methods based on DL algorithm can be divided into two categories according to whether the regional candidate network is adopted or not. The first stage is the two-

DOI: 10.4018/IJDCF.340934 *Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

stage target detection algorithm, the first stage is to generate candidate regions through the regional candidate network, and the second stage is to classify and regress the candidate regions (Xing et al., 2020). Literature (Zheng et al., 2021) put forward the histogram feature of gradient direction, combined with simple linear support vector machine as classifier, and achieved the best effect in the human detection algorithm at that time. Literature (Dao et al., 2022) proposes to use multi-camera pictures to detect the position, gesture and finger bending of hands, and stable detection can be realized through bone pictures, instead of data gloves, a contact device. The slow fusion model proposed in reference (Zimoch & Markowska-Kaczmar, 2021) uses 3D convolution and average pooling in its first three convolution layers, and achieved the best behavior detection effect at that time. Literature (Kohiyama & Yamashita, 2020) proposes a feature pyramid network for small target detection on the basis of Faster RCNN. Before using the feature pyramid structure, most DL-based detectors only detect at the top of the network.

Over the years, China has carried out a lot of research work in sensor network and monitoring, but the research combined with laboratory site safety monitoring is still in its infancy and exploration stage. Therefore, in order to manage hazardous chemicals in university laboratories scientifically, normatively and efficiently, it is particularly important to build an information management platform covering the whole process of purchase, use, storage and abandonment. In this paper, a multi-sensor integrated, flexible combination of functions, compact appearance, high precision and low cost laboratory dangerous operation behavior detection system is designed to improve the efficiency of laboratory safety management.Traditional methods for detecting hazardous work behaviors often rely on manually designed feature extraction methods, which often struggle to achieve ideal results in complex and dynamic environments. However, the emergence of deep learning algorithms provides new solutions to this problem. By constructing deep neural networks, we can automatically learn and extract high-level features from images, thereby achieving more accurate behavior detection.

In this article, we adopt a network structure of deep learning neural networks. This network structure can effectively capture spatial and temporal information in images. Firstly, deep learning is used to extract spatial features from raw images to identify the actions and postures of personnel in the laboratory. Then, deep learning is used to process temporal information to identify patterns of continuous actions and predict possible future behaviors. This network structure not only improves the accuracy of detection, but also enhances the real-time performance of the system.

## RESEARCH METHOD

### Detection Algorithm of Dangerous Operation Behavior in Laboratory

In recent years, the state has paid more and more attention to education and scientific research, continuously increased investment in university laboratory construction projects, and the laboratory has developed rapidly. With the growth of experimental demand, the types and quantities of hazardous chemicals used in laboratories are increasing. Each laboratory should formulate rules and regulations and operating procedures for the safety management of hazardous chemicals, record and monitor the whole process from purchase, collection, storage, use, recovery and disposal, strengthen the hardware conditions for the safety management of hazardous chemicals, and improve the technical ability and professional quality of safety officers. Because the management mode is limited by time and the number of managers, there is still room for further improvement in the current laboratory safety management.

The management of hazardous chemicals in universities is a complex and challenging task. Many laboratory experimenters in universities, especially scientific research experimenters, do not systematically record the detailed information of the purchasers, quantities, channels and varieties of hazardous chemicals in detail, resulting in missing or distorted purchase data. Universities should constantly explore the methods and modes of hazardous chemicals management, strengthen and standardize the safety management of hazardous chemicals in laboratories, reduce experimental accidents, ensure the personal safety of laboratory personnel, ensure the normal working order of teaching and scientific research, create a safe and good experimental environment, and ensure the teaching quality and scientific research output of laboratories.

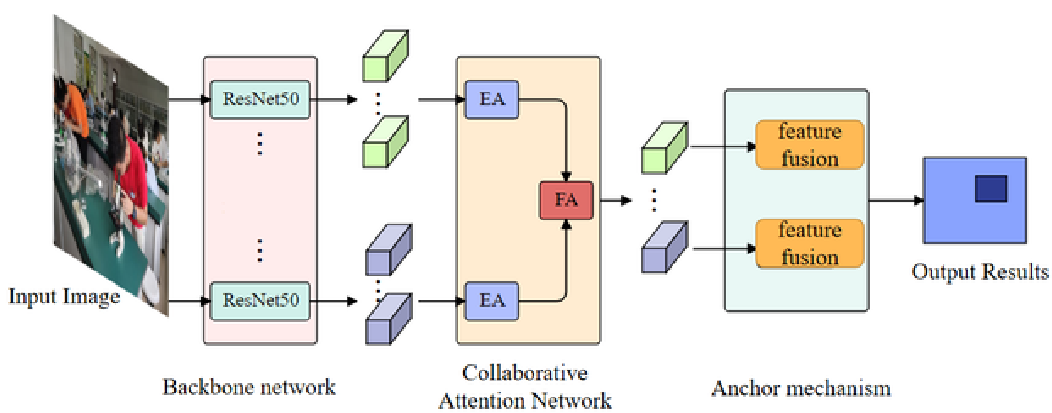
DL is a branch of machine learning and an important research field of artificial intelligence. DL is a neural network with deeper network level, more complicated calculation and more network parameters. Among them, CNN, as an important model of DL, plays a very important role in complex scenes such as unmanned driving, image recognition and classification, voice recognition and intelligent robots (Wang & He, 2021; Zhang et al., 2022). At present, the main steps of the target detection algorithm based on DL can be divided into the following steps: firstly, the features of the input image are extracted by the deep neural network, secondly, the regions with high probability of the target appearing are found on the feature map, and then these regions are classified, finally, the target is accurately located by the border regression.

CNN (Convolutional Neural Network) is a feedforward neural network, which is composed of convolution layer, pooling layer and full connection layer. This structure can make good use of the two-dimensional structure of input data. The model can also be trained using back-propagation algorithm. It is one of the most common structures in DL. RNN (Recurrent Neural Networks) is a kind of neural network with the ability of short-term memory (Shen et al., 2022; Sha et al., 2021). The output of the current sequence time is related to the previous sequence time and can process and predict sequence data. It is often used in natural language processing and other fields.

YOLO algorithm treats object detection as a regression problem. It is an end-to-end network. By inputting an image from the network, we can directly get its bounding box and category and the corresponding confidence probability. YOLO transforms the problem of target detection into a regression problem, and puts forward a simple method to speed up the detection (Hu et al., 2021; Qing et al., 2021). The bounding box and the category of the target can be predicted simultaneously by directly using a CNN, so YOLO can realize real-time video detection. In this paper, an algorithm framework of dangerous operation behavior detection in multi-task laboratory based on improved YOLOv5 structure is proposed, which uses end-to-end network to realize real-time head positioning, head segmentation and population regression. The flow chart of the algorithm is shown in Figure 1.

Firstly, the algorithm enhances, adaptively scales and adaptively anchors box computing on the input of YOLO network. Then convolution operation is carried out to strengthen the ability of network feature fusion (Lawal & Zhao, 2021). Finally, the GIoU_Loss function is used at the output to optimize the network parameters and accelerate the convergence of the model.

**Figure 1. Algorithm work flow chart**



There are a large number of convolution kernels in the three feature layers, but these convolution kernels make the image sample dimension very high, which leads to a lot of time consumption in calculation (Diana et al., 2019). The calculation method of parameter quantity of each convolution layer is as follows:

$$P = I_0 \times \left(1 + K_w \times K_h \times I_i\right) \tag{1}$$

where $P$ represents the number of total reference in each layer; $K_w, K_h$ is the width and height of convolution kernel respectively; $I_i$ is the number of input channels, and $I_0$ is the number of output channels of the upper convolution layer; Represents the number of convolution kernels in this layer; Adding 1 in parentheses means adding an offset to each convolution kernel.

For the judgment of dangerous operation behavior, it is necessary to process continuous video frames, and store an image and its detection results in the buffer area at certain intervals (Al-Antari et al., 2018; Yy, et al., 2020). When the buffer area is full, the occupied frame $Count_y$ and the unoccupied frame $Count_n$ in the buffer area are accumulated and counted respectively, and then the ratio is calculated. If it is greater than the departure judgment threshold $T_{leave} \in \left(1, B\right)$, it is output that there is someone, otherwise there is no one. As shown in formula (2), true stands for someone (Challita et al., 2018).

$$Res = true, if \left(\frac{Count_y}{Count_n}\right) > T_{leave} \tag{2}$$

The C3 structure of the original YOLOv5 network first maps the basic layer feature map into two parts in the channel dimension, and then maximizes the difference of the gradient information combination through the BottleNeck structure (Shin et al., 2022). Res2Block realizes fine particle feature fusion through a series of operations such as channel division, grouping convolution, inter-block fusion and channel splicing, and the specific process is shown in Formula (3).

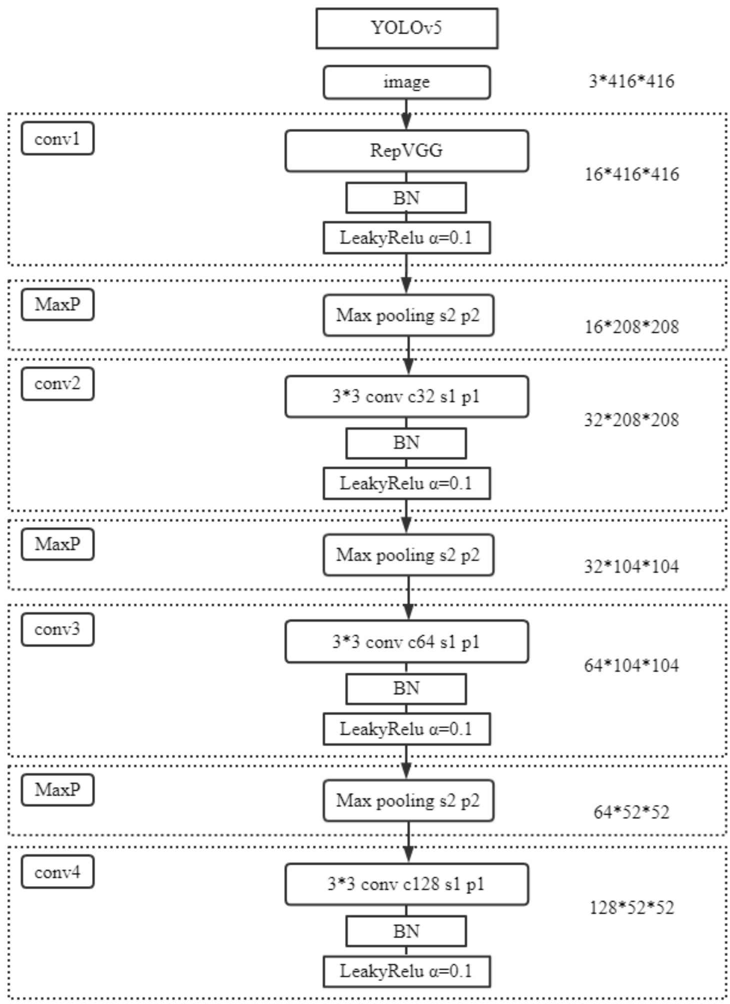
$$y_i = \begin{cases} x_i, & i = 1 \\ K_i\left(x_i + y_{i-1}\right), & 1 < i \le s \end{cases} \tag{3}$$

Among them, the input feature $x$ is divided into $s$ -block feature maps through channel division, $x_i$ represents the $i$ -block feature map, $K_i$ represents the convolution layer fused with the $i$ -block feature map, and $y_i$ represents the feature map obtained after fusing $x_i$ .

The same object may be displayed differently on the feature map due to different scales. This huge intra-class difference caused by different scales will reduce the detection performance of the model. This paper will explain this problem from the receptive field level of convolutional neural network. In this paper, a picture is randomly selected from the self-made data set, and the position and size of the laboratory students are marked. As shown in figure 2.

For the feature graph output by the middle layer of convolutional neural network, each point on the graph corresponds to the receptive field with a fixed size on the original graph, and conversely, the target on the original graph will be mapped to a fixed area on the feature graph of the middle layer. Different from the traditional fully connected network, each point of the fully connected network is related to the whole input. In order to cover pedestrians of different scales, one solution is to scale

**Figure 2. Different scales of laboratory students' behavior**

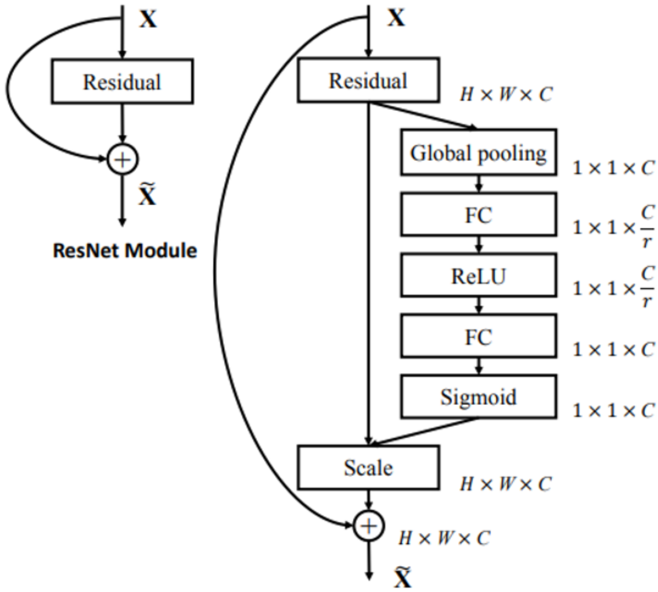



the image into several different scales, and then get the candidate frames of targets of different sizes, and then merge them for network training. Considering the above problems, this paper adopts the method of multi-layer candidate regions. As shown in Figure 3.

In this model, the strategy of generating candidate frames from multi-layer feature maps is adopted, and the candidate frames are trained by using anchor mechanism on feature maps of different network layers, so that the coverage of targets with different sizes can be completed without increasing the calculation amount.

Conv module of original YOLOv5 backbone network 3×3 adopts the design of direct connection of convolution kernel and activation function, which is often effective for feature extraction of pedestrian detection in non-dense scenes, but it is often difficult to extract features effectively for dense occlusion phenomenon. Therefore, RepVGG module is introduced and modified, and the original activation function ReLU is replaced by SiLU activation function (Smit et al., 2022; Hilloulin et al., 2022). RepVGG modules all adopt a four-layer multi-branch structure, which is composed of 3×3 convolution, 1×1 volume integral branch and residual difference branch of identity, in which the

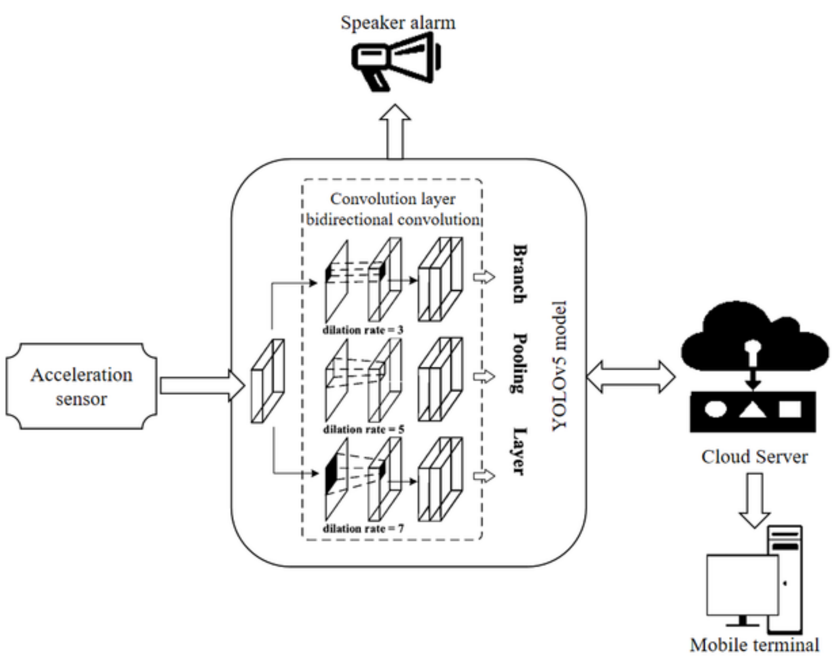
Figure 3. Schematic diagram of multi-layer candidate area acquisition




first layer is a down-sampling layer with a step size of 2. Wherein that improve backbone network structure is shown in Figure 4.

In this paper, Mosaic-9, an enhanced version of mosaic method, is adopted, that is, nine pictures are randomly cut, scaled and arranged to form a picture. Human attention is a unique brain signal

Figure 4. Improve the backbone network structure



processing mechanism of human vision (Healy et al., 2021). Attention mechanism is very similar to human visual attention. This paper introduces SENet module. The aggregation is followed by the excitation operation, which adopts a simple self-gating mechanism, which takes embedding as input and generates a set of corresponding weights for each channel. These weights are applied to u to generate the output of SE Block, which is then sent to the subsequent network.

The SENet attention mechanism is introduced into the bankbone of YOLOv5, and the Backbone module before and after introducing the SENet attention module can be represented as shown in Figure 5.

Then, YOLOv5 algorithm is used to obtain pedestrian area frame and Alphapose to estimate human posture. In this paper, regularization layer is added to standardize the obtained coordinate values, and the values in different ranges are mapped to specific intervals to make them evenly distributed, so that the network can learn and extract key information from the data more effectively (Wang et al., 2015 ; Nashat et al., 2010).

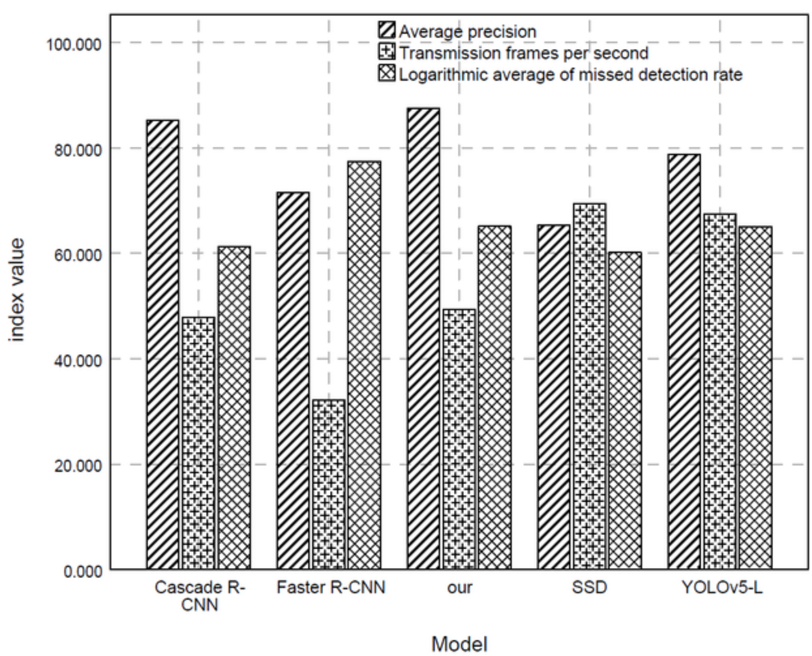
In this paper, the wrist is artificially raised to the middle of the shoulder and elbow, so as to simulate the posture of abnormal pedestrians using mobile phones. Take the lifting process of the left arm as an example, in which $P_1\left(x_1, y_1\right), P_2\left(x_2, y_2\right), P_1\left(x_3, y_3\right)$ represents the coordinate positions of shoulder, elbow and wrist respectively.

Write the length of $P_1, P$ as $l_{12}$:

$$l_{12} = \sqrt{\left(x_1 - x_2\right)^2 + \left(y_1 - y_2\right)^2} \tag{4}$$

Similarly, the length of $P_2, P_3$ is recorded as $l_{23}$:

**Figure 5. SE-backbone module**



$$l_{23} = \sqrt{\left(x_2 - x_3\right)^2 + \left(y_2 - y_3\right)^2} \tag{5}$$

The vertical distance from $P_2$ to line $l$ is denoted as $h_{12}$, and the coordinate $P_3^{'}\left(x_3^{'}, y_3^{'}\right)$ is calculated in the fitting process, where the definition of ordinate is shown in Formula (6):

$$y_3^{'} = \frac{y_1 + y_2}{2} \tag{6}$$

In the training process of DL, the loss value of the current model is obtained by calculating the difference between the real label in the image and the output result of the neural network, and the loss value reflects the difference between the output result and the real label. Usually, the loss of target detection network consists of three parts, namely, category loss, detection frame loss and confidence loss (Nguyen & Kim, 2013). This target detection task is single target detection without category loss, so the loss of this detection task is composed of head frame and confidence loss.

The loss of the accuracy of the reaction bounding box is marked as $loss_{box}$. The loss function of bounding box used in this paper is GIoU (Sha et al., 2021), and the calculation formula is as follows:

$$loss_{box} = 1 - \left(\frac{A \cap B}{A \cup B} - \frac{C - \left(A \cup B\right)}{C}\right) \tag{7}$$

Among them, $A$ is the target frame of the real image, $B$ is the object frame predicted by the network, and $C$ is the smallest bounding box of the $A$ and $B$ frames.

YOLOv5 uses GIoU_Loss as the loss function of Bounding box (Mira et al., 2012), as shown in Formula (5). An influence factor $v$ is added to the formula, and the ratio of the size coefficient of the prediction frame and the target frame is taken into account.

$$CIoU\_Loss = 1 - CIoU = 1 - \left(IoU \text{-} \frac{DIS\_2^2}{DIS\_C^2} - \frac{v^2}{\left(1 - IoU\right) + v}\right) \tag{8}$$

This GIoU_Loss, which increases the influencing factors, solves the problem of non-overlapping borders on the basis of IoU. Then YOLOv5 uses non-maximum suppression to filter many target frames.

In a given picture, firstly, we choose YOLOv5 algorithm to detect and obtain the area frame where pedestrians are in the picture, and then we independently estimate the coordinates of 17 joint points of each pedestrian by using Alphapose (Modenini et al., 2013). When the parameters of the four fully connected layers are set to (34, 256), (256, 512), (512, 128) and (128, 2) respectively, the detection accuracy reaches the highest.

### System Implementation

Because there are many laboratories in the school, and the environment and parameters required to be monitored in each laboratory are different, the developed system is required to have strong adaptability and expansibility. Therefore, we adopted the component-based software development technology in the development. The remote data acquisition and transmission based on Agent not only avoids the problem of heavy load of the remote monitoring host when the data flow of the monitoring system is large, but also improves the real-time and reliability of the system. More importantly, it reduces the frequent data exchange between the remote monitoring host and each monitoring point. To receive an IP packet, the router or host handles it as follows: first, search and match the routing table, preferably search the matching host first, and if a destination host with the same IP packet address is found, send the IP packet to the destination host.

The positioning of this system is to establish a laboratory intelligent monitoring system which is oriented to the public, can provide services for general laboratories and small and medium-sized enterprises' laboratories, and can meet the environmental information monitoring of basic laboratories. Functionally, it not only ensures the convenience of management based on the Web client, but also realizes the real-time data of the intelligent monitoring system in the laboratory, providing users with personalized and specific services. In order to inform the administrator of the danger in time, the embedded front-end machine should also have the ability to analyze the collected sensor data, give an alarm when it exceeds the threshold, and automatically cancel the alarm when it returns to the normal range. When the danger is found, the embedded front-end machine should also be able to automatically deal with the emergency to minimize the loss. At the same time, in order to deal with the danger better, the embedded front-end machine should also be able to accept the commands sent by the server and analyze and execute them.

In this paper, the DL-based laboratory dangerous operation behavior detection system uses the open source Darknet DL framework in the PC-side software development environment. After building the DL framework, it is necessary to install some acceleration libraries for DL, so that the model can be accelerated by GPU in training and testing, so as to improve the real-time performance of the whole system. YOLO's pre-training network is trained on ImageNet 1000-class data set. The pre-training network uses the first 20 convolution layers plus an average pooling layer as shown in the above figure, and finally adds a fully connected layer. Then the pre-trained weights of the first 20 convolution layers are applied to the detection as initialization weights, and the remaining 4 convolution layers and 2 fully connected layers are added at the same time. Only the last fully connected layer in the network uses the linear activation function.

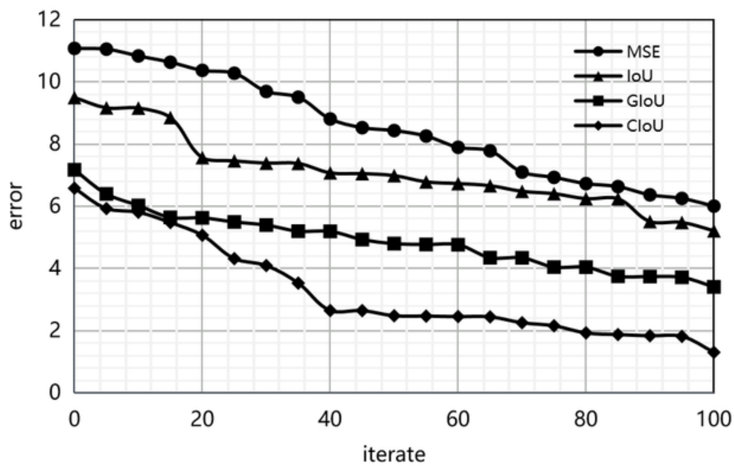
The hardware part of the environmental multi-parameter wireless monitoring system in the laboratory is mainly the data acquisition terminal, and the whole terminal mainly includes four parts: sensor detection part, microcontroller control processing part, GPRS wireless module transmission part and system power module part. Sensors are widely used in the monitoring system to collect signals, so that people can get the current situation in real time. Sensor information acquisition module is the source part of wireless security system, which mainly includes three aspects of signal acquisition. As shown in Figure 6, it is a system block diagram:

The collected data are input to neural network for training after wavelet denoising and dimensionality reduction. Then identify the detected data, give an alarm if there is a fall, and synchronize the information to the cloud server, and the user will be notified by the Web application SMS.

The embedded front-end machine is mainly responsible for collecting real-time data from the laboratory, processing it and uploading it to the monitoring center. Monitoring center. The monitoring center grasps the situation of each monitored point in real time. The administrator can check the environmental conditions of the laboratory through the remote monitoring terminal. The monitoring center is a computer with a fixed IP and connected to the Internet through an Ethernet card. Through the monitoring software, the monitoring center can grasp the situation of each monitored point in real time, so as to make data statistics and analysis and respond and deal with the situation. It can monitor the environmental conditions of each laboratory very intuitively.

In the actual use process, university experimenters are often tired of recording the use information of dangerous chemicals, and some experimenters don't even have the habit of recording the use information, which leads to blank data and difficult traceability. Therefore, the voice function

**Figure 6. System chart**




should be added to the usage module. As long as the experimenter uses voice, he can transmit the usage information to the mobile phone, and then connect it with the information intelligent system through the mobile phone, and then import the information into the system. When there is a fire in the laboratory, the module detects the flame, and combined with the output signals of the temperature sensing and smoke sensing module, the main control chip judges whether there is a fire. If there is a fire, the buzzer will be started in time and early warning information will be sent to the designated mobile phone number to realize the remote early warning function, so that the laboratory administrator can take further measures.

## EXPERIMENTAL ANALYSIS

In this paper, the algorithm is tested based on the constructed behavior recognition data set. Ten videos of different scenes are used in the experiment, all of which are recorded by TP-LINK webcam. These shooting angles clearly capture people in the laboratory from different directions, and have different laboratory environments and lighting conditions. The data set is divided into 80% training set and 20% test set. The training samples of the data set are rich and diverse, and the scale is large enough to train a model with good effect.

The improved YOLOv5 target detection model is implemented based on the PyTorch framework (Hu et al., 2022). The experimental hardware environment is: Intel(R) Core (TM) i7 processor, 16G DDRL memory, NVIDIA GTX 2080Ti GPU; The software environment of the experiment includes Ubuntu 64-bit operating system environment, Python3.6 and PyTorch1.0DL framework.

Improved YOLOv5 network model adopts the method of obtaining candidate regions from multi-layer feature layers. In the experiment, this paper combines the candidate regions obtained from different feature layers, then trains them separately, and finally compares the performance of the model. As shown in Table 1.

In this paper, the candidate regions obtained from different feature layers are combined in a progressive way, in which “fusion” means that the feature layer with this magnification combines the features of high-level and middle-level. The acquisition of multi-level candidate regions and the fusion of features have a great influence on the performance of the model, especially in terms of accuracy, which shows that target detection from feature maps with different magnification can improve the accuracy of detection. In the task of dangerous operation behavior detection, due to the different scale

**Table 1. Feature fusion comparison**

| Magnifying Power | Recall (%) | Accuracy (%) | Time (ms) |
|---|---|---|---|
| 32 | 65.871 | 56.742 | 43.366 |
| 32+16 | 66.871 | 56.742 | 45.286 |
| 32+16 (Fusion) | 67.063 | 59.368 | 46.561 |
| 32+16 (Fusion) +8 | 75.798 | 69.201 | 47.181 |
| 32+16 (Fusion) +8 (Fusion) | 79.798 | 82.458 | 48.226 |

distribution of human body, obtaining candidate regions on feature maps with different magnification can better cover human bodies with different scales, thus improving the detection accuracy.

In real life, most of the videos taken by surveillance videos are normal, and a robust abnormal behavior detection method should have a low false alarm rate for normal behaviors. Deep network can automatically learn and predict the location of abnormal behavior in video. In order to achieve this goal, the network should score high abnormal scores for abnormal video clips during the training iteration. This means that this section with high abnormal score is the time axis position where abnormal behavior occurs. Because of the complexity of these real anomalies, it is impossible to complete the anomaly detection task well only by using normal behavior data.

As shown in Table 2, the influence of adding header information on the experimental results of detecting abnormal behavior is demonstrated. Among them, after adding head information on the basis of using arm information, the accuracy is improved in both fitting data sets and real data sets, and the accuracy in fitting data sets is always higher than that in real data sets. Compared with 14.923% when only hand information is used, the probability of missing alarm is reduced to 7.155% after adding head information, which also proves the effectiveness of adding head information.

In order to verify the effect and performance of YOLOv5-MPL proposed in this paper, the algorithm is compared with other mainstream target detection algorithms, and the results are shown in Figure 7.

The experimental results show that the proposed method is the best algorithm among the above algorithms, and it does have good performance. In order to solve the overlapping problem caused by the original data set in the laboratory scene, the positioning information of the receptive field is strengthened without increasing the network parameters basically, and the detection ability of the model for the human body in the laboratory scene is enhanced.

The last stage of human behavior recognition is classification, and different classification algorithms have different effects on human behavior classification. In order to improve the accuracy of human behavior classification, it is very important to select the appropriate classification algorithm. If the training data is approximately separable, a classifier can be learned to maximize the soft interval, which is also called soft interval support vector machine. If the training data is inseparable, kernel function can be introduced on the basis of learning a classifier to maximize the soft interval to solve

**Table 2. Influence of arm and head information on experimental results**

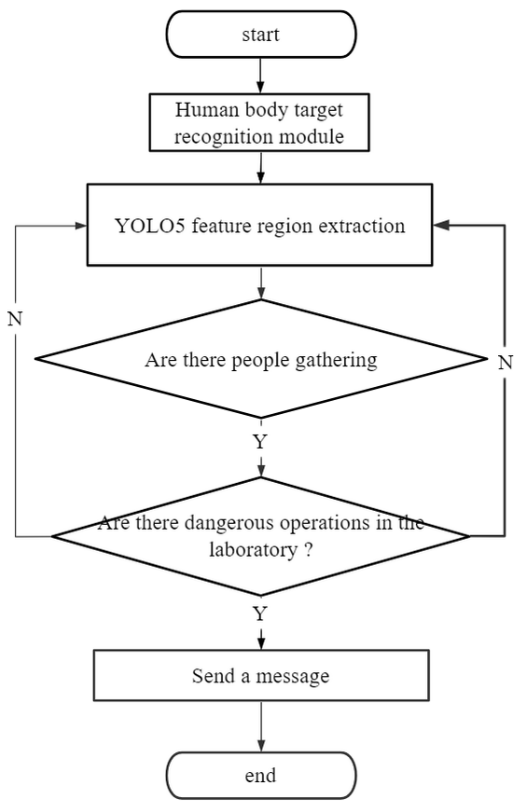
| Structure | Accuracy (Fitting)/% | Accuracy (True)/% | Missed Alarm Probability (True)/% | False alarm Probability (True)/% | Recall Rate (True)/% |
|---|---|---|---|---|---|
| arm | 94.181 | 94.984 | 14.923 | 15.547 | 88.436 |
| Arm+head | 96.308 | 97.765 | 7.155 | 6.201 | 97.313 |

Figure 7. Contrast experimental results

the problem of data inseparability. In this paper, only six-axis data are extracted, that is, three-axis acceleration data and three-axis gyroscope data. Because CNN is used to train the data obtained by the sensor, the input data of CNN is based on images, so it is necessary to convert the data obtained by the sensor into images.

The effects of different loss functions on network performance are analyzed experimentally, and the experimental results are shown in Figure 8:

Figure 8. Error curve of different loss function training



From the objective of minimizing the loss function, the MSE loss function is the sum of Euclidean distances that minimize the four coordinates of $\left(x, y, w, h\right)$, and has no obvious directionality. Compared with MSE loss function, IoU and GIoU loss functions are more in line with the performance evaluation index of target detection, so their final effect is better than MSE loss function.

Because CIoU loss can directly optimize the distance between the center point of the predicted bounding box and the center point of the real bounding box, the convergence speed is much faster than that of MSE, IoU and GIoU loss functions, and the final error has reached a lower level.

Fig. 9 is an image captured in the simulation experiment. The red unlabeled box is the behavior of missing laboratory students manually marked after the screenshot, and the blue box is the false detection. In the experiment, the model runs stably, although the processing speed does not reach the rate of 30 frames per second, it can basically meet the real-time effect through some other methods, such as frame extraction processing or the assistance of other sensors.

The methods based on deep learning are significantly better than traditional methods in terms of accuracy and real-time performance. Specifically, for different types of hazardous homework behaviors, deep learning methods have achieved better accuracy than traditional methods. In addition, deep learning methods also demonstrate excellent real-time performance. To test the robustness of the system, we conducted tests in different laboratories. The number of personnel, lighting conditions, background, etc. in these laboratories vary. The experimental results show that regardless of the environment, the behavior detection system based on deep learning can maintain high accuracy and real-time performance. This proves that the system has strong robustness and adaptability.

We also conducted adaptive learning experiments to test the system's ability to recognize new dangerous behaviors. In the experiment, we simulated new hazardous work behaviors and observed the system's response. The experimental results show that the system can quickly adapt to new situations and accurately identify new dangerous behaviors. This is due to the strong adaptive ability of deep learning algorithms.

In summary, through comparative experiments, robustness experiments, and adaptive learning experiments, we have fully verified the superiority and effectiveness of the laboratory hazardous work behavior detection system based on deep learning algorithms. This system not only has high accuracy

**Figure 9. Example of simulation experiment results**

and real-time performance, but also strong robustness and adaptability. This provides strong technical support for laboratory safety management and helps to reduce the incidence of safety accidents.

## CONCLUSION

There are many kinds of hazardous chemicals in university laboratories, with scattered storage locations and dense users, and they are flammable, explosive, toxic, infectious and corrosive, and some highly hazardous chemicals even have highly toxic characteristics. Over the years, China has carried out a lot of research work in sensor network and monitoring, but the research combined with laboratory site safety monitoring is still in its infancy and exploration stage. In this paper, a multi-sensor integrated, flexible combination of functions, compact appearance, high precision and low cost laboratory dangerous operation behavior detection system is designed to improve the efficiency of laboratory safety management. In this paper, an algorithm framework of dangerous operation behavior detection in multi-task laboratory based on improved YOLOv5 structure is proposed, which uses end-to-end network to realize real-time head positioning, head segmentation and population regression. The research results show that this method is the best among the comparison algorithms, and it does have good performance. In order to solve the overlapping problem caused by the original data set in the laboratory scene, the positioning information of the receptive field is strengthened without increasing the network parameters basically, and the detection ability of the model for the human body in the laboratory scene is enhanced.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

## FUNDING STATEMENT

This work is supported by Liaoning Provincial Department of Education's Basic Research Project for Universities [grant number JYTMS20230711] and Liaoning Province Science and Technology Plan Joint Program (Fund) Project [Grant No. 2023JH2/101700009].

## REFERENCES

Challita, U., Dong, L., & Saad, W. (2018). Proactive resource management for lte in unlicensed spectrum: A deep learning perspective. *IEEE Transactions on Wireless Communications*, *17*(99), 1–1. doi:10.1109/TWC.2018.2829773

Chen, B., Liu, Y., & Sun, K. (2021). Research on object detection method based on ff-yolo for complex scenes. *IEEE Access : Practical Innovations, Open Solutions*, *2021*(99), 1–1. doi:10.1109/ACCESS.2021.3119624

Dao, B. T., Nguyen, T. V., Pham, H. H., & Nguyen, H. Q. (2022). Phase recognition in contrast-enhanced ct scans based on deep learning and random sampling. *Medical Physics*, *2022*(7), 49. doi:10.1002/mp.15551 PMID:35428990

Dewi, C., Chen, R. C., Liu, Y. T., Jiang, X., & Hartomo, K. D. (2021). Yolo v4 for advanced traffic sign recognition with synthetic training data generated by various gan. *IEEE Access : Practical Innovations, Open Solutions*, *2021*(99), 1–1. doi:10.1109/ACCESS.2021.3094201

Diana, S., & Damira, P., & Bagheri, M. (2019). In-yolo: Real-time detection of outdoor high voltage insulators using uav imaging. *IEEE Transactions on Power Delivery, 35*(3), 1599-1601.

Gómez, A., & Muñoz, A. (2023). Deep learning-based attack detection and classification in Android devices. *Electronics (Basel)*, *12*(15), 3253. doi:10.3390/electronics12153253

Healy, E. W., Tan, K., Johnson, E. M., & Wang, D. L. (2021). An effectively causal deep learning algorithm to increase intelligibility in untrained noises for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *149*(6), 3943–3953. doi:10.1121/10.0005089 PMID:34241481

Hilloulin, B., Bekrine, I., Schmitt, E., & Loukili, A. (2022). Open:ource deep learning-based air(oids detection algorithm for concrete microscopic images. *Journal of Microscopy*, *286*(2), 179–184. doi:10.1111/jmi.13098 PMID:35286722

Hu, D., Zhu, J., Liu, J., Wang, J., & Zhang, X. (2022). Gesture recognition based on modified yolov5s. *IET Image Processing*, *2022*(8), 16.

Kohiyama, M., Oka, K., & Yamashita, T. (2020). Detection method of unlearned pattern using support vector machine in damage classification based on deep neural network. *Structural Control and Health Monitoring*, *202*(4), e2552. doi:10.1002/stc.2552

Lawal, O. M., & Zhao, H. (2021). Yolofig detection model development using deep learning. *IET Image Processing*, *15*(13), 3071–3079. doi:10.1049/ipr2.12293

Li, H., Deng, L., Yang, C., Liu, J., & Gu, Z. (2021). Enhanced yolo v3 tiny network for real-time ship detection from visual image. *IEEE Access : Practical Innovations, Open Solutions*, *2021*(99), 1–1. doi:10.1109/ACCESS.2021.3053956

Li, H. J., Wang, Y., Wu, L. Y., Liu, Z. P., Chen, L., & Zhang, X. S. (2012). Community structure detection based on potts model and network's spectral characterization. *Europhysics Letters*, *97*(4), 48005–48010. doi:10.1209/0295-5075/97/48005

Mira, A., Bhattacharyya, D. K., Saharia, S., Mira, A., Bhattacharyya, D. K., & Saharia, S. (2012). Rodha: Robust outlier detection using hybrid approach. *American Journal of Epidemiology*, *2*(5), 129–140.

Modenini, A., Rusek, F., & Colavolpe, G. (2013). Optimal transmit filters for isi channels under channel shortening detection. *IEEE Transactions on Communications*, *61*(12), 4997–5005. doi:10.1109/TCOMM.2013.110813.130385

Muñoz, A., Palomo, E. J., & Jerez-Calero, A. (2019). Use of an ANN to value MTF and melatonin effect on ADHD affected children. *IEEE Access, 7*. https://doi.org/.127254-127264

Nashat, D., Jiang, X., & Kameyama, M. (2010). Group testing based detection of web service ddos attackers. *IEICE Transactions on Communications, B*(5), 1113-1121.

Nguyen, T., & Kim, J. M. (2013). Multistage optical smoke detection approach for smoke alarm systems. *Optical Engineering (Redondo Beach, Calif.)*, *52*(5), 7001. doi:10.1117/1.OE.52.5.057001

Ophoff, T., Puttemans, S., Kalogirou, V., Robin, J. P., & Goedemé, T. (2020). Vehicle and vessel detection on satellite imagery: A comparative study on single-shot detectors. *Remote Sensing (Basel)*, *12*(7), 1217. doi:10.3390/rs12071217

Qing, Y., Liu, W., Feng, L., & Gao, W. (2021). Improved yolo network for free-angle remote sensing target detection. *Remote Sensing (Basel)*, *13*(11), 2171. doi:10.3390/rs13112171

Sha, Z., Feng, H., Rui, X., & Zeng, Z. (2021). Pig tracking utilizing fiber optic distributed vibration sensor and yolo. *Journal of Lightwave Technology*, *2021*(99), 1–1. doi:10.1109/JLT.2021.3073225

Shen, H., Dong, Z., Yan, Y., Fan, R., Jiang, Y., & Chen, Z. et al.. (2022). Building roof extraction fro astil echo images applying osa-yolov5s. *Applied Optics*, *2022*(11), 61.

Shin, H., Kim, J. K., Choo, Y. J., Choi, G. S., & Min, C. C. (2022). Prediction of motor outcome of stroke patients using a deep learning algorithm with brain mri as input data. *European Neurology*, *85*(6), 460–466. doi:10.1159/000525222 PMID:35738236

Smit, V., Thüsen, J. H., Cohen, D., Hondelink, L. M., Hüyük, M., & Postmus, P. E. et al.. (2022). Development and validation of a supervised deep learning algorithm for automated whole-slide programmed death-ligand 1 tumour proportion score assessment in non-small cell lung cancer. *Histopathology*, *80*(4), 635–647. doi:10.1111/his.14571 PMID:34786761

Wang, D., & He, D. (2021). Channel pruned yolo v5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, *210*(6), 271–281. doi:10.1016/j.biosystemseng.2021.08.015

Wang, X., Zeldovich, N., Kaashoek, M. F., & Solar-Lezama, A. (2015). A differential approach to undefined behavior detection. *ACM Transactions on Computer Systems*, *33*(1), 1–29. doi:10.1145/2699678

Xing, J., Yu, M., Wang, S., Zhang, Y., & Ding, Y. (2020). Automated fraudulent phone call recognition through deep learning. *Wireless Communications and Mobile Computing*, *2020*(2), 1–9. doi:10.1155/2020/8853468

Xu, D., & Wu, Y. (2020). Improved yolo-v3 with densenet for multi-scale remote sensing target detection. *Sensors (Basel)*, *20*(15), 4276. doi:10.3390/s20154276 PMID:32751868

Yy, A., Kanaji, A. S., Matsuda, A. T., Oshikiri, A. T., Nakamura, A. T., & Suzuki, A. S. et al.. (2020). Automated surgical instrument detection from laparoscopic gastrectomy video images using an open source convolutional neural network platform. *Journal of the American College of Surgeons*, *230*(5), 725–732. doi:10.1016/j.jamcollsurg.2020.01.037 PMID:32156655

Zhang, Y., Xu, W., Yang, S., Xu, Y., & Yu, X. (2022). Improved yolox detection algorithm for contraband in x-ray images. *Applied Optics*, *61*(21), 61. doi:10.1364/AO.461627 PMID:36256244

Zheng, J., Yu, K., & Wu, X. (2021). A deep model based on lure and similarity for adaptive clickbait detection. *Knowledge-Based Systems*, *214*(5–6), 106714. doi:10.1016/j.knosys.2020.106714

Zimoch, M., & Markowska-Kaczmar, U. (2021). Human flow recognition using deep networks and vision methods. *Engineering Applications of Artificial Intelligence*, *104*(1), 104346. doi:10.1016/j.engappai.2021.104346