Journal of Southeast University (English Edition) Vol. 39, No. 4, pp. 372 - 383 Dec. 2023 ISSN 1003—7985

Efficient fire-smoke detection and evacuation simulation from buildings based on YOLOX-Swin

Xu Zhao Dai Tianqi

(School of Civil Engineering, Southeast University, Nanjing 211189, China)

Abstract: To achieve efficient emergency response and evacuation from building fires, the possibility of applying object detection technology to building fire emergency management is investigated. An application of object detection algorithms in the early warning stage of fire is proposed by combining a transformer, convolutional neural network, and lightweight attention mechanism module (namely convolutional block attention module) to extract the local and global features of flames and smoke, thereby improving the accuracy of the object detection algorithm and achieving fast localization of fire occurrence. An improved ant colony algorithm for path searching is proposed by improving the heuristic function and pheromone evaporation coefficient. A grid map model is developed in a case, and the effectiveness of the proposed method is verified through simulation and emulation by considering positioning information. The results show that compared with the YOLOX algorithm, the YOLOX-Swin model improves the average accuracy by 1.5%. The improved ant colony algorithm reduces the search range of the traditional ant colony algorithm, improves the convergence speed of the model, and avoids the problem of getting trapped in local optimum solutions. By integrating early warning of fire and personnel evacuation, a comprehensive building fire emergency management plan is developed.

Key words: computer vision; self-attention; ant colony algorithm; fire dynamics simulator (FDS)

DOI: 10.3969/j.issn.1003-7985.2023.04.006

Fire is a symbol of progress in human civilization, which provides warmth and light to humanity. However, uncontrolled fire can yield huge disasters. Fire, as used here, refers to catastrophic combustion that is uncontrolled during its spread, causing certain damage to human life and property. Although the destructive power of fires and the resulting degree of damage may not be as large as those of natural disasters such as floods and earthquakes, their frequency of occurrence is beyond comparison with other natural disasters. Therefore, compared with other natural catastrophes, fires are significantly more destructive over the long run, causing enormous losses to human society in terms of property and life and are a catastrophic issue currently faced by people all over the world. The severity and diversity of building fires have undergone significant changes, and in recent years, they have become an increasingly important topic of concern. According to statistics, in the past two decades, a total of 86 million fires have occurred worldwide, resulting in over one million deaths and an average economic loss of up to $857.9 billion per year, representing 1% of the world gross domestic product[1].

Building fires are characterized bythe rapid spread of smoke, difficulty in controlling and extinguishing the fire, and personnel evacuation[2]. To minimize casualties, rapid and effective fire warning and evacuation path guidance are particularly important. Early flame and smoke detection methods use smoke sensors, temperature sensors, and traditional smoke detection algorithms based on artificial features like the color, edge, texture, and other characteristics of smoke; however, these methods suffer from several limitations, such as high false alarm rates, poor real-time performance, and insufficient generalization ability, which cannot meet actual industrial needs.

Following the development of deep learning (DL) algorithms, DL-based flame-smoke detection models can balance the relationship between real-time performance and accuracy in a very efficient manner. Smoke detection algorithms can be divided into classification network methods (image and video classifications) and detection network methods based on the different neural networks used. He et al.[3] proposed the ResNet residual network model, which could train deeper networks and solve the problem of gradient disappearance in deep networks. Their approach was to load the results of shallow networks onto the results of deeper networks. To study smoke detection in foggy weather, Khan et al.[4] created a foggy-smoke dataset and fine-tuned the VGG16 network for four classification tasks: normal and foggy smoke detections.

In addition to the classification-based methods, Zhang et al.[5] used an object detection network called a faster

Received 2023-04-01, **Revised** 2023-06-10.

Biographies: Xu Zhao (1982—), male, Ph. D., associate professor, xuzhao@seu.edu.cn.

Foundation items: The National Natural Science Foundation (No. 72071043), the Natural Science Foundation of Jiangsu Province (BK20201280), Humanities and Social Science Fund of Ministry of Education (20YJAZH114).

Citation: Xu Zhao, Dai Tianqi. Efficient fire-smoke detection and evacuation simulation from buildings based on YOLOX-Swin[J]. Journal of Southeast University (English Edition), 2023, 39(4): 372 - 383. DOI: 10.3969/j.issn.1003-7985.2023.04.006.

region-based convolutional neural network (CNN) to detect wildfires in the wild and used synthetic methods to obtain data labeled with target boxes to avoid manual labeling. Megvii Technology[6] proposed a high-performance detection algorithm, namely, YOLOX, which was based on the anchor-free network structure that relied on YOLOv5. YOLOX adopted a decoupled head and the leading label assignment strategy named SimOTA, which effectively accelerated the convergence of the model and greatly reduced its training time. Moreover, Vision Transformer (ViT)[7] is a pioneering work that introduces transformers in natural language processing tasks into a visual field. ViT processes images in a block-like form similar to text and achieves excellent results in image classification problems. However, applying the transformer to the detection of smoke and flames in fires is confronted with two main challenges. Firstly, the entity changes and shapes of the smoke and flames are large and not fixed; thus, the performance of this method may not be as good as required. Secondly, the detection of fire scenes requires accuracy and speed; however, because of high image resolution, the transformer based on the global self-attention mechanism requires a large computational complexity and is not sufficiently fast.

In the field of fire protection, researchers often use computer simulation techniques to simulate fire development and personnel evacuation behavior scenarios. However, traditional software simulation methods suffer from many limitations, such as inaccurate simulation environments and significant errors compared with reality, which lead to inaccurate results. Therefore, these limitations restrict the usefulness of fire evacuation simulation software for research on fire escape planning. Moreover, building information modeling (BIM) is a new technology that has emerged in recent years and has allowed building digital technology to reach higher levels[8]. BIM has become a hotspot in the field of engineering construction. With the advancement of this technology, combining fire simulation and personnel evacuation behavior becomes possible, in which accurate building structure, model information, and environmental parameters are provided by BIM technology, making the research results more closely aligned with reality. Furthermore, BIM technology can quickly establish a building model in the event of a disaster and coordinate with other numerical simulation software. At this level, fire dynamics simulator (FDS) is a tool that is developed by the National Institute of Standards and Technology Building Fire Research Laboratory, which consists of a fire-driven fluid flow model using computational fluid dynamics[9]. Zhao et al.[10] used FDS software to study the characteristics of smoke movement during a fire in a high-rise building staircase, including the temperature and pressure distribution on a vertical surface, the location of the neutral plane in the chimney effect, and other related phenomena. In recent years, ant colony optimization (ACO) has achieved important results in the field of crowd safety evacuation[11]. By combining BIM with the ant colony algorithms, the Revit platform is used to construct building models, and the *PyroSim* dynamic fire simulation software is used to obtain smoke temperature and smoke concentration data in a fire environment, which provides the base data for parameter values to the ant colony algorithm. This integration enables visualization of crowd evacuation, facilitating a better understanding of the evacuation process in fire scenarios. To improve the global search capability of the algorithm, Wang[12] proposed an improved ACO with a bidirectional search strategy to optimize fire evacuation routes. Xia et al.[13] proposed an ancient village fire escape path planning method based on improved ACO, which introduced two factors, namely, path length and path reliability, and added heuristic information to improve search efficiency.

The flame smoke-detection model based on YOLOX and Swin Transformer (ST)[14] with sliding window operations and hierarchical design can achieve accurate detection of multiscale targets after satisfying the requirements of real-time detection. To achieve an efficient emergency response to fires, in addition to real-time and accurate fire warning devices, research on fire spread mechanisms and personnel evacuation path planning is required. Therefore, the *PyroSim* fire numerical simulation software is used to obtain the numerical values such as environmental temperature and volume fraction of each major evacuation node under fire scenes, and an emergency response plan for fires in buildings is developed by combining the optimized ACO.

1 Proposed Integrated Framework

The current study proposes a deep neural network-based early warning system for building fires and smoke. By utilizing an object detection model, the system can detect flames and smoke and identify the source of the fire by analyzing the location of the camera that captures it. The system then employs pre-simulation fire dynamics software to acquire data regarding the environmental temperature and concentration of harmful gases within the building. Subsequently, the system designs secure evacuation pathways for those trapped using the ant colony algorithm that is optimized for building fire environments. The best evacuation route is obtained, which aids and directs trapped individuals in their escape.

1.1 DL-based object detection of fire smoke

Following the development of object detection, the YOLO series is pursuing an optimal trade-off between speed and accuracy in real-time applications. Over the past two years, anchor-free detectors, advanced label as-

signment strategies, and end-to-end detectors have realized major achievements in object detection academia. However, they are not yet integrated into the YOLO family. Because YOLOv4 and YOLOv5 may be a little over-optimized for anchor-based pipelines, YOLOX is improved based on the YOLOv3 and provides these recent advancements to the YOLO series with an experienced optimization. Compared with previous YOLOv1 to YOLOv5, the accuracy of YOLOX in object detection has greatly improved[15]. Therefore, the present study chooses YOLOX as the basic algorithm for flame smoke detection, and improvements are made to it.

1.1.1 YOLOX network model structure

YOLOX algorithm is a high-performance one-stage object detection network, which was proposed by the MEGVII Research Institute in 2021. It cleverly applies excellent progress in object detection, such as decoupled heads, data augmentation, and anchor-free methods, to YOLO. The network structure is shown in Fig. 1, which consists of three parts, namely, the CSPDarknet53 network used as the backbone feature extraction network, a path-aggregation network (PANet) used to enhance the feature extraction network in the neck part, and three decoupled heads used in the prediction part[16]. The backbone part extracts the shallow features from an input image to obtain three feature layers, which will be used by the neck part to extract the deep features. Finally, each of the three decoupled heads is used for the object box detection to obtain the search results. The following improvements are primarily made by the YOLOX algorithm: it separates the decoupling heads and separately implements classification and regression[17].

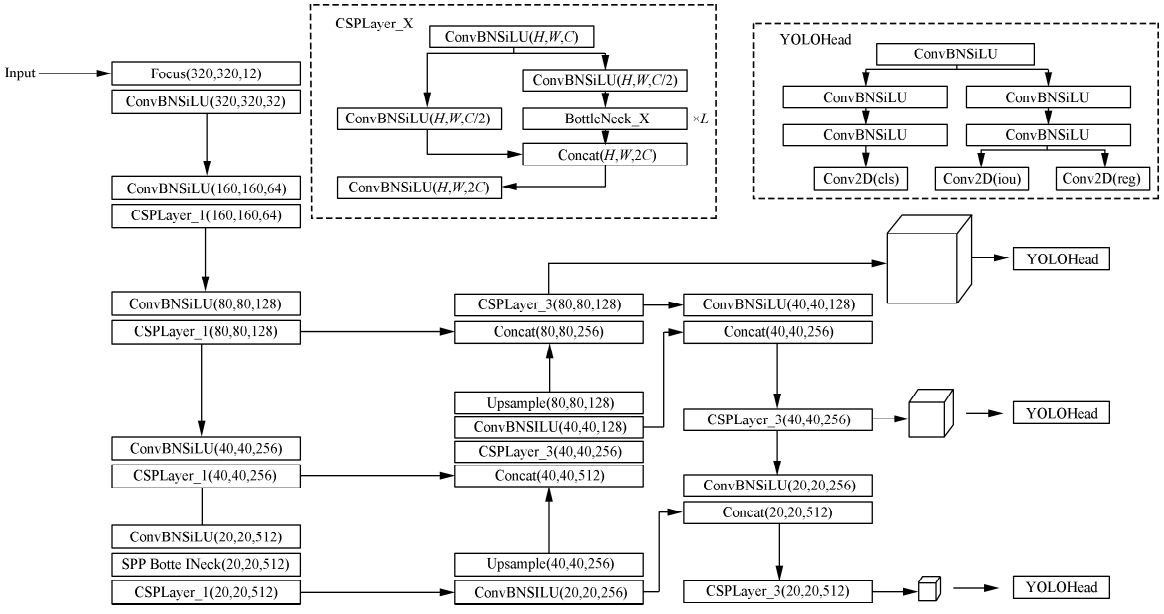


Fig. 1 Network architecture of YOLOX

1.1.2 YOLOX-Swin

In the YOLO series, YOLOX exhibits excellent detection performance and has exceeded the accuracy of YOLOv3 to YOLOv5. However, the backbone of YOLOX is based on traditional CNN. The convolution kernel can effectively capture the local feature information in the image, but it can easily ignore the global feature information. To better detect small flames, a lightweight attention-mechanism module, namely the convolutional block attention module (CBAM), is added to the feature pyramid network layer of YOLOX, which can effectively enhance the detection ability of small target objects[18-19]. Moreover, the transformer structure with a self-attention mechanism can well obtain global feature information of an image, which helps the network greatly improve the detection effect of large objects in the images. In the flame-smoke detection process, smoke is usually large, whereas flame is generally small. Therefore, this paper proposes a YOLOX-Swin model for flame-smoke detection by combining the features of the convolution kernel and transformer structure, and the model structure is shown in Fig. 2.

The partial CSPLayer structure in the backbone of the YOLOX feature extraction network has been replaced with CSPSwin Block to enhance the model perception of global information features. Fig. 3 shows that for a given input feature $\boldsymbol{X} \in \mathbf{R}^{2C \times H \times W}$, CSPSwin Block uses the Conv_BN_SiLU basic convolution module to reduce the dimensionality of the input feature to C and then performs a transpose operation to obtain $\boldsymbol{X}_{\text{patch}} \in \mathbf{R}^{C \times H \times W}$. Next, $\boldsymbol{X}_{\text{patch}}$ is fed into the ST module to obtain the global feature information. By implementing the transpose process, $\boldsymbol{X}_{\text{trans}} \in \mathbf{R}^{C \times H \times W}$ can be obtained with global feature information.

CBAM is an attention mechanism that combines channel and spatial attention. Fig. 4 shows that, given a feature

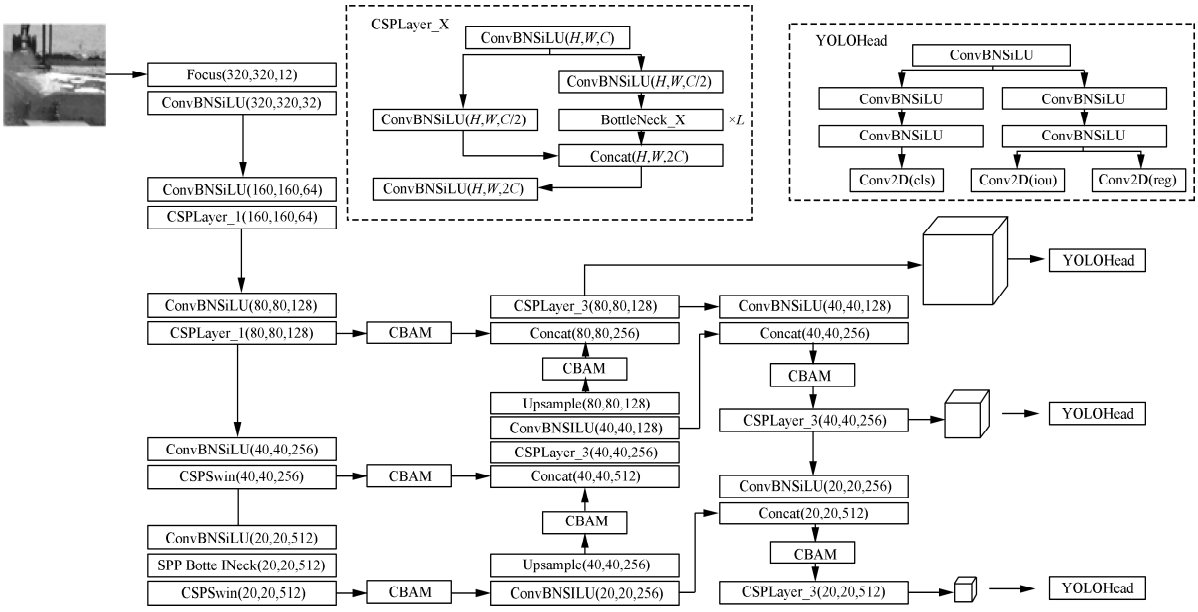


Fig. 2 Network architecture of YOLOX-Swin

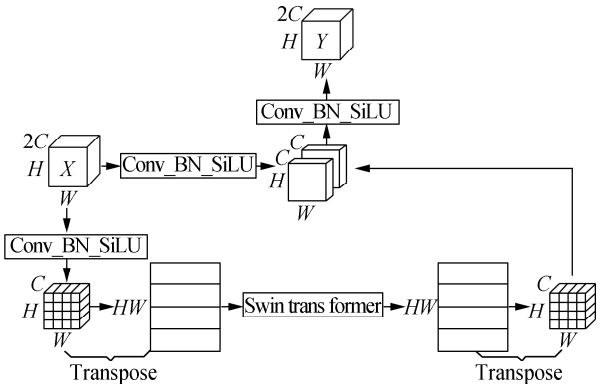


Fig. 3 Network architecture of CSPSwin block

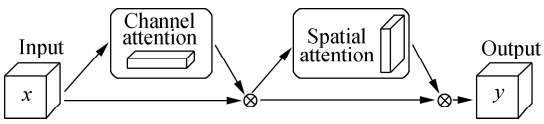


Fig. 4 CBAM

map, CBAM can sequentially generate attention feature maps in both channel and spatial dimensions. The two feature maps are then multiplied with the original input feature map for adaptive feature refinement to produce the final feature map. As a lightweight attention mechanism module, CBAM can be embedded in the backbone network of an object detection model without greatly increasing the number of parameters, thereby improving the model performance.

1.2 Emergency evacuation model using ACO

The personnel evacuation path planning problem in building fires involves finding the safest path for people to escape in the shortest time. The evacuation time depends on not only the actual length of the path but also the evacuation speed of the personnel. In a fire scenario, the evacuation speed of the personnel is mainly affected by factors such as temperature, fire products (such as carbon monoxide), and crowd density. The influence of fire products and crowd density on the evacuation speed of the personnel can be quantified as temperature influence coefficient $f_1(T)$ and carbon monoxide concentration influence coefficient $f_2(\theta_{co})$. By combining the actual length of the path and the influence coefficient of the evacuation speed of the personnel, the equivalent path length can be obtained. The shorter the equivalent path length is, the shorter the required evacuation time.

1.2.1 ACO

ACO is a bioinspired algorithm first proposed in 1992 by Italian researcher Dorigo M. It is inspired by the foraging behavior of ants in nature. ACO provides the advantages of positive feedback, heuristic search, and distributed computing and is widely used to solve optimization problems in path planning with feasible solutions[20].

During their movement, ants leave a substance on their path to communicate information, which is called a pheromone. Ants can perceive pheromones during their movement, which can influence their direction of movement. Under initial conditions, i. e., at $t=0$, m ants are placed at starting point S, and ant $k(k=1,2,\ldots,n)$ moves from S to endpoint E. When an ant encounters a path selection, it is more likely to choose the path with a higher concentration of pheromone. The probability of ant k moving from node i to node j is given by

$$p_{ij}^{(k)}(t)=\begin{cases}\dfrac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}}{\sum\limits_{s\in N_k}[\tau_{is}(t)]^{\alpha}[\eta_{is}(t)]^{\beta}} & j\in N_k\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

where $\tau_{ij}(t)$ represents the pheromone concentration between nodes i and j; $\eta_{ij}(t)$ denotes the heuristic function; d_{ij} denotes the Euclidean distance between nodes i and j; α stands for the importance factor of the pheromone, which is usually set to 0.5; β denotes the importance factor of the heuristic function, which is usually set to 0.5; and N_k denotes the set of nodes available for the next step. When each ant makes a path selection or walks along a complete path, it needs to update the pheromones. The rule for updating the pheromones is expressed as follows:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) \tag{2}$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t) \tag{3}$$

where ρ is the pheromone evaporation coefficient, which has a value range from zero to one; $\Delta\tau_{ij}(t)$ is the total change in pheromones from node i to node j; $\Delta\tau_{ij}^{k}(t)$ is the number of pheromones released by ant k on the path from node i to node j. The calculation of $\Delta\tau_{ij}^{k}(t)$ uses the ant cycle model, and its expression is shown as

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} \dfrac{Q}{L_k} & \text{passing through segment } ij \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where Q represents the constant of the pheromone; L_k denotes the total distance traveled by ant k in the current cycle.

1.2.2 ACO improvement

This study uses the grid method to establish an environmental model of the first floor of a large building, as shown in Fig. 5. The grid method is a global path planning method that divides the environment into free space and obstacles using equally sized rectangular grids, which offers the advantages of easy encoding and computation[21]. Black grids represent obstacles such as walls and pillars in the building, and each obstacle grid can have up to eight neighboring grids, four of which are corner grids. If the neighboring grids of an obstacle corner grid are non-obstacle grids, then the corner grid is called a vertex grid, as shown in the gray grids in Fig. 5. Personnel can freely move in the white and gray grids in the figure. Generally, buildings have large construction areas and complex internal structures, and using their grid maps for path searching can result in too many alternative points and high computational complexity. To reduce the computational complexity, the gray grids shown in Fig. 5 are extracted as candidate points for ACO. Then, the connectivity among the vertex grids needs to be determined. When no obstacle or danger grid exists (temperature exceeding 120 ℃ or concentration reaching 2.5×10^{-3}) between two vertex grids, no connectivity exists between the two grids.

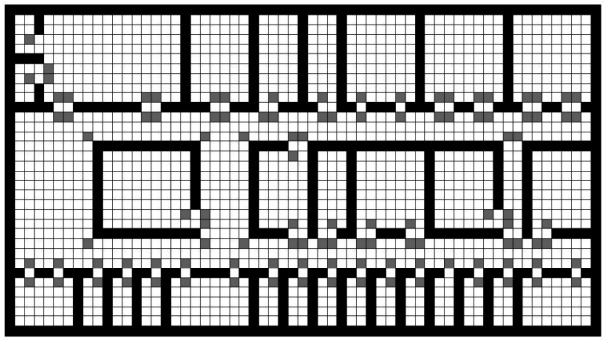


Fig. 5 Grid map

The grid index is used to represent the position index of a grid. For a grid map with a size of R rows and C columns, the relationship between grid index i and coordinates of the grid center point (x_i, y_i) is expressed as

$$\left.\begin{aligned} x_i &= \mathrm{mod}(i, C) - 0.5 \\ y_i &= R - \mathrm{ceil}\left(\frac{i}{C}\right) + 0.5 \end{aligned}\right\} \tag{5}$$

where mod ( ) denotes the modulo function and ceil ( ) denotes the ceiling function.

When a fire occurs, the temperaturerises due to the convective heat of the smoke and the radiant heat of the flame. An appropriate body-sensed temperature (30 to 60 ℃) will have a positive effect on the evacuation of personnel. However, as the body-sensed temperature continues to rise to 60 to 120 ℃, the evacuation speed of the personnel will gradually decrease until they lose their ability to move. At this point, high temperature exerts a negative effect on the evacuation of personnel. Temperature-effect coefficient $f_1(T)$ on the evacuation of personnel can be calculated as

$$f_1(T) = \begin{cases} 1 & T < T_{s1} \\ \dfrac{(v_{max} - 1.2)\left(\dfrac{T - T_{s1}}{T_{s2} - T_{s1}}\right)}{1.2} + 1 & T_{s1} \leqslant T < T_{s2} \\ \dfrac{v_{max}}{1.2}\left[1 - \left(\dfrac{T - T_{s2}}{T_d - T_{s2}}\right)^2\right] & T_{s2} \leqslant T < T_d \end{cases} \tag{6}$$

where T is the temperature of the environment at the scene of fire; T_{s1} is the temperature at which a person feels uncomfortable (generally considered to be 30 ℃); T_{s2} is the temperature that causes harm to the human body (generally considered to be 60 ℃); T_d is the lethal temperature (generally considered to be 120 ℃); V_{max} stands for the maximum evacuation speed of personnel (usually considered to be 4 m/s).

The personnel evacuation speed is affected by not only the body-sensed temperature at the fire scene but also the concentration of toxic and harmful gases released during the combustion of combustibles. Among them, carbon

monoxide is the main gas that causes casualties to the evacuating personnel during fires. When the volume fraction of carbon monoxide in the smoke is less than 0.1%, the effect on the personnel evacuation can be ignored. However, when the volume fraction reaches 0.25%, the evacuating personnel may suffer serious harm, and they can experience difficulty in moving. The effect coefficient of the carbon monoxide volume fraction on the evacuation of personnel at the fire scene can be calculated as follows:

$$f_2(\theta_{co}) = \begin{cases} 1 & \theta_{co} < 0.1\% \\ 1 - (0.212\,5 + 1.788\theta_{co}) & 0.1\% \leqslant \theta_{co} < 0.25\% \\ 0 & \theta_{co} \geqslant 0.25\% \end{cases} \tag{7}$$

where θ_{co} denotes the volume fraction of carbon monoxide.

When the abovementioned two fire-effect coefficients are considered, the shortest geometric path at the fire scene may not necessarily be the path with the shortest evacuation time for the personnel. This study replaces the Euclidean distance in the heuristic function with an equivalent distance, which is calculated as follows:

$$D_{ij}(t) = \frac{L_{ij}\delta_{if}}{M_{ij}(t)} \tag{8}$$

$$M_{ij}(t) = f_{ij}(T) f_{ij}(\theta_{co}) \tag{9}$$

where $D_{ij}(t)$ denotes the equivalent distance between nodes i and j at time t; δ_{ij} stands for the degree of difficulty of passage between nodes i and j, usually set as 1.5; L_{ij} represents the Euclidean distance between nodes i and j; $M_{ij}(t)$ denotes the coefficient of the effect of a high-temperature fire environment and carbon monoxide volume fraction of the fire products on the personnel evacuation.

Because of the absence of large differences in the concentration of pheromones among the different paths and lack of orientation toward the target point in the early stage of the algorithm, a high degree of randomness occurs in the ant path selection, resulting in a slow convergence rate of the algorithm and a decrease in the search efficiency and ability to find the optimal path. The use of equivalent distance instead of the Euclidean distance can effectively reduce the randomness of the ant path selection. The equivalent distance is expressed as

$$\eta_{ij} = \frac{1}{D_{ij} + d_{jE}} \tag{10}$$

where D_{ij} denotes the equivalent distance between nodes i and j; d_{jE} stands for the Euclidean distance between node j and endpoint E.

The pheromoneevaporation factor ρ largely affects the convergence speed and search efficiency of the algorithm. Therefore, the size of the evaporation factor ρ is dynamically adjusted. Initially, a high pheromone evaporation factor is set, and the pheromone evaporation factor is then reduced as the number of iterations increases to update the pheromones and improve the convergence and global search capabilities of the algorithm. The specific method is expressed as

$$\rho(t) = \mathrm{Min}[0.95\rho(t-1),\ \rho_{\min}] \tag{11}$$

2 Results and Discussion

2.1 Fire-smoke recognition based on YOLOX-Swin

2.1.1 Acquisition of flame smoke image dataset and data augmentation

This study used two different types of data sources, including an image that was open data on the web and a video, which were transformed into a video image sequence, and some highly similar pictures were removed. Finally, 2 500 images of fire and fireworks were retained. The dataset contained 3 552 flame and 1 347 smoke targets. The LabelImg dataset annotation software was used to label the flame and smoke in the images.

Data augmentation can generate more data from limited datasets, increase the diversity of the dataset, and enhance the robustness of the model. Because the dataset used in the present study was relatively small, the methods of Mosaic and Mixup data augmentation could effectively expand the training dataset.

Mosaic is animage data augmentation method that combines four different images by randomly cropping and splicing them together to create a larger image. The positions and sizes of the crops and splices are random, which makes the generated images more diverse and complex. Mosaic data augmentation helps expand limited datasets and improve the performance of object detection models, which makes the trained models more adaptable to complex scenes and environments. In this study, Mosaic data augmentation was used in the first 85% of the epochs, and the last 15% of the epochs were used for closing. In the later stages of training, when the learning rate decreased, disabling Mosaic could reduce the batch size of the training data, thereby fine-tuning the model and improving its performance.

Mixup is an image data augmentation method that generates new images by blending the target objects from two different images. Specifically, for two randomly selected images, their pixel values are weighted and fused according to a random ratio, generating a new image. The corresponding target labels are then linearly interpolated according to the same ratio. Mixup data augmentation is commonly used in image classification and object detection tasks, especially on small datasets, to improve model performance by increasing the diversity and size of the

dataset. The advantage of Mixup data augmentation is that it can increase the diversity of the dataset and improve the generalization ability of the model without increasing any computational cost. Additionally, Mixup can reduce the risk of overfitting and enhance the robustness of the model.

2.1.2 Model training

Because of the modification of the backbone feature extraction network in this study, the pretrained network on the ImageNet dataset could not be used during the training phase; thus, the model needed to be trained from scratch. During the training, the dataset was divided into a 9∶1 ratio for the training and validation sets as well as for a separate testing set. Within the (training + validation) set, the dataset was divided into a 9∶1 ratio for the training and validation sets. The input image resolution was 640 × 640, and the stochastic gradient-descent optimizer and cosine annealing learning rate were used for the training. The batch size was 16, the initial learning rate was 0.01, and the model was trained for 400 iterations.

The training and testing were performed using the platform of an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20 GHz, 128 GB DDR4 RAM, CUDA Toolkit 11.6, and GPU NVIDIA Tesla T4. The operating system was the CentOS 7 workstation.

2.1.3 Evaluation

In the object detection classification problem, the objects to be classified were positive samples, and the other objects, including the background, were negative samples. IOU represents the ratio of the intersection over the union between the predicted bounding box and the actual labeled bounding box. The classification of the predicted results into positive and negative samples was based on the IOU threshold. For a given object, the predicted position by the model was usually represented as a rectangular box, also known as a bounding box. The actual labeled position was also a rectangular box. The IOU threshold is usually set to 0.5. When IOU ≥ 0.5, the predicted box was considered a true positive; otherwise, it was considered a false positive.

Recall refers to the proportion of positive samples that were correctly identified by the model among all positive samples. It is one of the most important metrics for evaluating a model's performance in recognizing targets because it measures the proportion of targets that are missed by the model during recognition. Precision refers to the proportion of predicted positive samples that were correctly predicted as positive by the model among all samples predicted as positive. Precision is one of the metrics used to measure the accuracy of model predictions. The specific equations are

$$R_e = \frac{T_P}{T_P + F_N} \tag{12}$$

$$P_r = \frac{T_P}{T_P + F_P} \tag{13}$$

Average precision (AP) represents the area under the precision-recall curve. High AP indicated that the object detection model maintained a good precision rate while keeping a high recall rate. For multiclass classification problems, mean AP (mAP) is commonly used to characterize the performance of object detection models.

The training loss of YOLOX-Swin is shown in Fig. 6, which shows that the model gradually converged after 250 rounds of training according to the loss curve. Because of the cessation of the Mosaic data augmentation at the epoch of 340, a significant decrease in the training loss occurred. Overfitting did not occur during the training process.

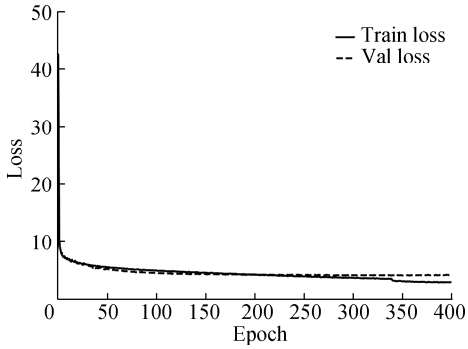


Fig. 6 Loss curve

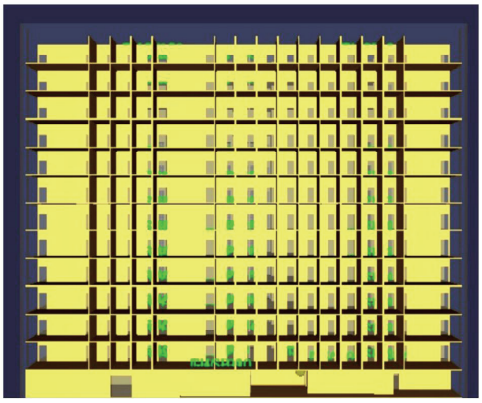
In this study, a series of ablation experiments was meticulously designed to systematically evaluate the effect of YOLOX improvements on the network performance utilizing the lightweight network structure "-s." Three comprehensive experiments were conducted to rigorously validate the effects of the improvements on the performance of the YOLOX model, where all training parameters were consistently maintained. The model performance detection outcomes are listed in Table 1, where "√" denotes the strategy employed in the improved model and "×" signifies the strategy that was not incorporated in the improved model. Meticulous analysis of the experimental results revealed that Improvement 1 involved the modification of the original CSPLayer module to the CSPSwin module, which effectively fused both CNN and sliding window transformer modules to enable extraction of local and global features from the image in the backbone feature extraction network. This modification resulted in an increment of 7×10^5 in the number of parameters and a noteworthy improvement of 0.9% in mAP. Furthermore, Improvement 2 entailed the addition of the CBAM (channel and spatial) lightweight attention module based on Improvement 1, which greatly enhanced the model ability to extract informative features in both channels and spatial dimensions. Despite a slight increment of 1×10^5 in the number of parameters, this improvement led to a considerable improvement of 0.6% in mAP.

Table 1 Comparison of the algorithm performance

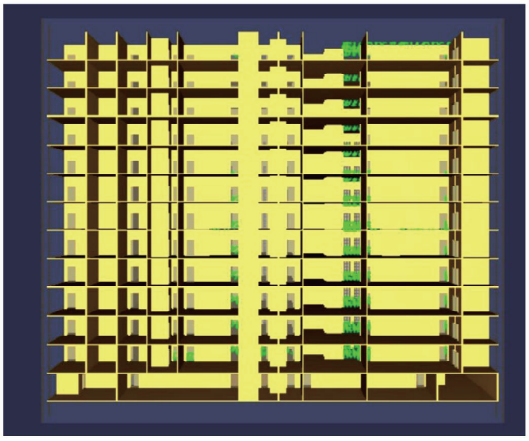
Method	CSPSwin	CBAM	Parameters/10^6	mAP/%
YOLOX	×	×	9.0	68.8
Improvement 1	√	×	9.7	69.7
Improvement 2	√	√	9.8	70.3

2.2 Numerical simulation of fire smoke

The biomedical building under investigation in this study was limited to the main building, which spanned a single floor with a total construction area of approximately 1 800 m^2 and a height of 57.3 m. The detailed floor plan of the building is shown in Fig. 7, which indicates the location of staircases on the east and west sides, along with three passenger elevators on the west side and one fire elevator on the east side. Fig. 7 also shows the back and front elevation views of the building, providing a comprehensive visual representation of its structural characteristics.



(a)



(b)

Fig. 7 Front and rear elevation views. (a) Front-elevation view; (b) Rear elevation view

2.2.1 Model establishment

This study presents a numerical simulation study of fire-smoke dynamics during the construction phase of a building, considering the limitations of software modeling and the nonoccurrence of actual fires. To simplify the model, the underground structure, structural beams, and columns were omitted. The walls were assumed to have a uniform thickness of 200 mm, whereas the slabs were set at a consistent thickness of 150 mm. The installation of external glass curtain walls was not completed on any floor, and only shear walls were constructed indoors. Filling walls and lightweight partition walls were not included in the model. The walls and slabs were assumed to be constructed of nonflammable concrete materials, and wooden templates were piled up in the room where the fire occurred, in which the wooden template material remained on the walls as per the main construction phase. Doors and windows were not installed during this phase, and in the model, windows and doors were represented as reserved holes to allow for automatic ventilation. No fire facilities such as smoke exhaust, fire hydrant, or automatic sprinkler systems were assumed to exist in the building. The mesh size used in the simulation was set at 0.5 m × 0.5 m × 0.5 m, resulting in a total of 682 000 grids.

2.2.2 Fire scene and prerequisites

As a combustible material commonly used in concrete pouring during the construction period of buildings, wood formwork material is often the cause of fires that occur during the main structure phase. Therefore, in this study, wood formwork material was selected as the combustible material for simulation. The fire room was surrounded by wood formwork material, and piles of wood formwork were present in the room.

The fire source size was 1 m × 1 m, and the central coordinate was (36, −12.5, 1). The fire rapidly developed in a square form, and the maximum heat release rate of the fire source was 3 296 kW/m^2. The prerequisites were as follows. 1) The combustion reaction of *n*-heptane, which was more suitable for actual conditions, was selected. 2) The ignition material was yellow pine from the software material library, and the default parameters of the software were used. 3) The airflow caused by the movement of construction personnel was not considered.

The boundary conditions were as follows: 1) No temperature difference exists between the building indoors and outdoors. The initial environmental temperature was set at 20 ℃, and the atmospheric pressure was standard atmospheric pressure. 2) The gas inside the building and the smoke generated by the fire were both ideal gases. 3) The simulation time of the model was set to 600 s. 4) Only a vent was set at the top of the model at $z = 54$ m.

2.2.3 Results

During the construction phase of the main building structure, the absence of installed exterior glass curtain walls and interior doors and windows created favorable conditions for the fire, including ample oxygen supply, stable combustion environment, and efficient ventilation, as shown in Fig. 8. As the fire progressed, the resulting

smoke rapidly propagated within the building space. At T = 60 s, the smoke primarily extended outside of the building on the north-facing facade and quickly ascended toward the upper floors due to the lack of external glass curtain walls. At T = 90 s, the smoke began to spread toward the third and fourth floors while continuing to rise. At T = 140 s, the smoke reached the ninth floor, with some infiltrating the seventh, eighth, and ninth floors while the rest continued ascending. At T = 180 s, the smoke diffused to the top floor, with some also extending to the twelfth floor, while the remaining smoke was discharged through the ventilation openings. At this stage, the smoke started to disperse along the corridor in both directions through the reserved holes on the third and fourth floors, which filled the upper part of the corridor and spread toward the upper floors on the east-facing facade. At T = 240 s, the smoke had already occupied the corridors on the third, fourth, seventh, eighth, and ninth floors and began to spread toward the upper part of the building on both the east-and west-facing facades. Finally, at T = 500 s, the smoke had fully spread throughout the entire building.

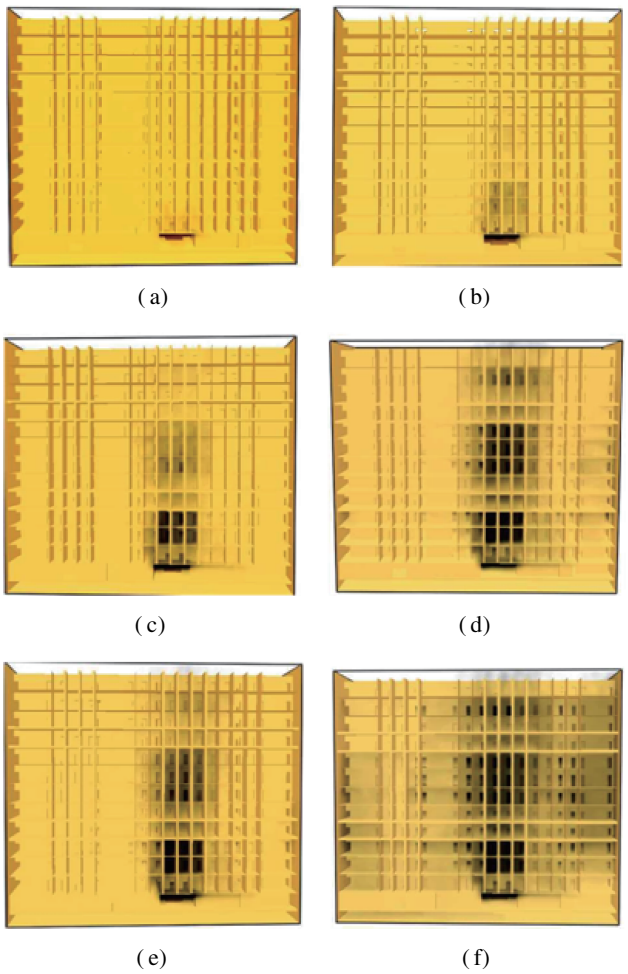


Fig. 8 Fire-smoke diffusion on the front facade at different times. (a) T = 60 s; (b) T = 90 s; (c) T = 140 s; (d) T = 180 s; (e) T = 240 s; (f) T = 500 s

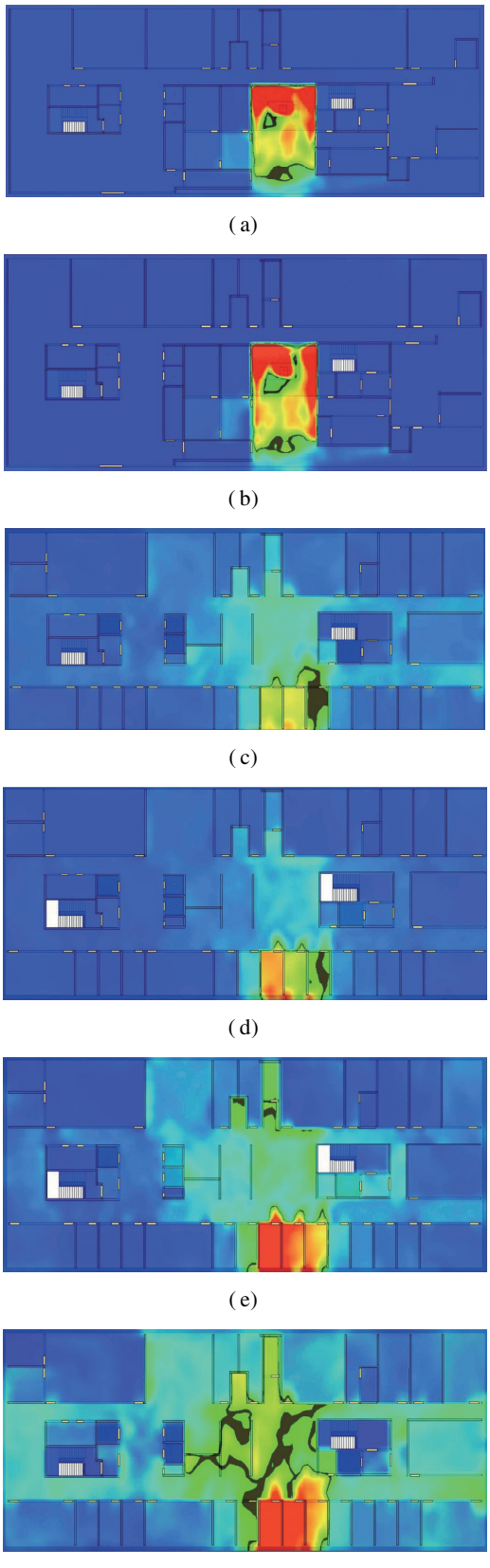
The temperature profile of the smoke at a distance of 2 m above the floor on the first, third, and fourth floors at T = 400 s when the smoke temperature has reached a steady state was analyzed in this study. The selection of a 2-m height was comparable to the average height of a person's head during normal walking. On the first floor, the high-temperature smoke was primarily concentrated close to the fire source with a maximum temperature of 98.5 ℃, which posed potential risks to human health. Hence, during the evacuation procedures, rescue personnel should avoid traversing areas close to the fire source. As the smoke propagated to the third and fourth floors along the building facade, the temperature distribution of smoke on these floors was similar. The graphical representation showed that the high-temperature smoke was predominantly concentrated in four small rooms, and the smoke subsequently spread to the corridors and adjacent rooms through the reserved holes, which led to an increase in temperature in the central hall. Notably, because of the extensive spread of the smoke, the smoke temperature on the fourth floor was higher than that on the third floor.

The distribution of carbon monoxide concentration at the height of 2 m above the floor on the first, third, and fourth floors at T = 400 s when the carbon monoxide concentration has achieved a steady state was analyzed in this study. Fig. 9 shows that the distribution pattern of carbon monoxide closely resembled that of the smoke temperature on these three floors. On the first floor, areas with higher volume fraction of carbon monoxide were primarily concentrated close to the fire source, reaching a maximum of 0.25%. Carbon monoxide spread through the reserved holes on the first floor and along the vertical surface to the adjacent rooms and upper part of the building. Upon reaching the third and fourth floors, carbon monoxide infiltrated through four small rooms and rapidly disseminated throughout the entire floor via the holes. Notably, Fig. 9 shows that the carbon monoxide concentration on the fourth floor was very much higher than that on the third floor, and the spread of carbon monoxide was more extensive, practically filling the entire floor. Consequently, personnel on the fourth floor faced a greater risk of carbon monoxide exposure during evacuation. On the third and fourth floors, carbon monoxide predominantly accumulated in the four small rooms and central hall. Thus, to ensure safety during rescue operations for trapped individuals, these areas should be avoided as much as possible.

2.3 Simulation of fire evacuation of personnel

2.3.1 Establishment of the simulation environment

The present study used the main building of a biomedical facility as the engineering context for the investigation. Specifically, the first floor of the building, which



(a) (b) (c) (d) (e) (f)

Fig. 9 Distribution of flue gas temperature and carbon monoxide at $T = 400$ s. (a) Temperature distribution of smoke on the ground floor; (b) Carbon monoxide distribution on the ground floor; (c) Temperature distribution of smoke on the third floor; (d) Carbon monoxide distribution on the third floor; (e) Temperature distribution of smoke on the fourth floor; (f) Carbon monoxide distribution on the fourth floor

encompassed an approximate construction area of 1 800 m^2, was considered, assuming the presence of a single designated safe exit. The scenario of a fire outbreak in Optical Clean Room 2 on the first floor was postulated as the focal point of analysis. To facilitate the simulation experiment, a grid map with dimensions of 27 × 68 was constructed where each grid cell measures 1 m. The simulation was implemented using the Matlab software. Notably, because of the sensitivity of the ACO algorithm to its parameter settings, a cautious approach was taken to ensure reliable results. Prior to conducting the simulation experiment, the initialization of each parameter was meticulously controlled by altering only one parameter at a time while keeping the others constant in a series of 10 simulation runs. Subsequently, the results were averaged to obtain the optimal parameter values for subsequent experiments. This systematic approach guaranteed the robustness and accuracy of the simulation outcomes. Optimal parameters $\alpha = 1$, $\beta = 1.4$, $Q = 12$, $m = 50$, and $\rho = 0.9$ were obtained in this manner.

α denotes the importance factor of the pheromone. β is the importance factor of the heuristic function. Q represents a constant indicating the strength of the pheromone. m denotes the number of ants. ρ is the pheromone evaporation coefficient.

2.3.2 Results

In this study, the experiment was conducted using two groups: one simulating personnel evacuation before the occurrence of fire and the other simulating evacuation after changes in the site environment due to fire. The starting point for personnel evacuation was set at (66.5, 3.5), whereas the coordinates of the exit were (1.5, 13.5). The improved ACO algorithm was employed to identify the shortest path for personnel evacuation toward a safe exit before the fire outbreak. The simulation results are shown in Fig. 10, where the dotted line indicates the shortest evacuation path, START represents the starting point, and GOAL represents the endpoint.

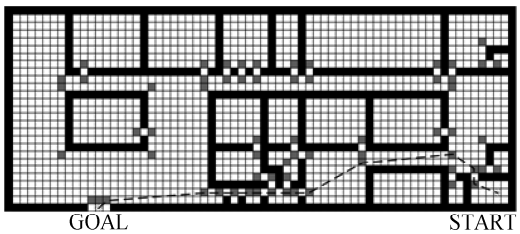


Fig. 10 Simulation results before the fire

Furthermore, information such as the environment temperature and carbon monoxide volume fraction at the fire scene was obtained, and the red area in Fig. 10 represents the dangerous grid that was impassable due to the occurrence of the fire. The starting and ending coordinates for evacuation remained the same as those before the fire, and the results of personnel evacuation path planning are

shown in Fig. 11, which evidently illustrate that the growth of fire and its spread rendered nodes (48.5, 7.5) and (48.5, 8.5) as dangerous grids because of the elevated environmental temperature and carbon monoxide volume fraction. The algorithm intelligently avoided the path with the shortest Euclidean distance and instead selected a path with a relatively shorter equivalent distance, circumventing the fire area and imminent spread area. This process ensured a more favorable evacuation path that prioritized personnel safety.

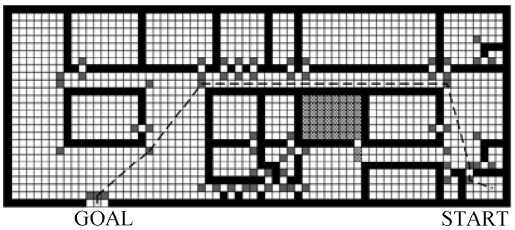


Fig. 11 Simulation results during the fire

3 Conclusions

1) We implement a fire-smoke detection method based on the YOLOX-Swin object detection model. At an IOU threshold of 0.5, the achieved mAP reaches 70.3%. Notably, compared with YOLOX, our improved model demonstrates a 1.5% improvement in mAP while maintaining a similar parameter volume, indicating superior performance in flame-smoke detection.

2) We performed a numerical simulation of fire and smoke in construction sites using the *PyroSim* software. A model based on the biomedical-building plan is established using the *PyroSim* software. This study analyzes the probability of fires caused by different combustible materials at various construction stages and simulates the combustion of template wood materials during the main structural construction stage, which is most susceptible to fires. The obtained smoke temperature and carbon monoxide volume fraction information provides critical data for path planning for the safe evacuation of trapped individuals.

3) We propose improved ACO for fire evacuation path planning of trapped personnel. On the basis of basic ACO, this study enhances the heuristic function and pheromone update coefficient by incorporating the characteristics of building fires. This process enables rapid planning of safe evacuation paths for trapped personnel based on the on-site fire environment once a fire occurs in a building.

References

[1] Kodur V, Kumar P, Rafi M M. Fire hazard in buildings: Review, assessment and strategies for improving fire safety[J]. *PSU Research Review*, 2019, **4**(1): 1 − 23. DOI: 10.1108/PRR-12-2018-0033.

[2] Guanquan C, Jinhua S. Quantitative assessment of building fire risk to life safety[J]. *Risk Analysis*, 2008, **28**(3): 615 − 625. DOI: 10.1111/j.1539-6924.2008.01048.x.

[3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//*IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Las Vegas, NV, USA, 2016: 770 − 778. DOI: 10.1109/CVPR.2016.90.

[4] Khan S, Muhammad K, Mumtaz S, et al. Energy-efficient deep CNN for smoke detection in foggy IoT environment[J]. *IEEE Internet of Things Journal*, 2019, **6**(6): 9237 − 9245. DOI: 10.1109/JIOT.2019.2896120.

[5] Zhang Q, Lin G, Zhang Y, et al. Wild land forest fire smoke detection based on faster R-CNN using synthetic smoke images[J]. *Procedia Engineering*, 2018, 211: 441 − 446. DOI: 10.1016/j.proeng.2017.12.034.

[6] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO series in 2021[J]. (2021-07-18)[2022-12-15]. http://arXiv.org/pdf/2107.08430v2.pdf.

[7] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16 × 16 words: Transformers for image recognition at scale[J]. (2020-10-22)[2022-12-15]. http://arXiv.org/pdf/2010.11929.pdf.

[8] Porwal A, Hewage K N. Building information modeling (BIM) partnering framework for public construction projects[J]. *Automation in Construction*, 2013, **31**: 204 − 214. DOI: 10.1016/j.autcon.2012.12.004.

[9] Sun L, Podila K, Chen Q, et al. Computational fluid dynamics modeling of fire and human evacuation for nuclear applications[J]. *Journal of Nuclear Engineering and Radiation Science*, 2020, **6**(1): 011112. DOI: 10.1115/1.4044531.

[10] Zhao G, Beji T, Merci B. Study of FDS simulations of buoyant fire-induced smoke movement in a high-rise building stairwell[J]. *Fire Safety Journal*, 2017, **91**: 276 − 283. DOI: 10.1016/j.firesaf.2017.04.005.

[11] Yang X, Yang X, Li Y, et al. Obstacle avoidance in the improved social force model based on ant colony optimization during pedestrian evacuation[J]. *Physica A: Statistical Mechanics and Its Applications*, 2021, **583**: 126256. DOI: 10.1016/j.physa.2021.126256.

[12] Wang J. Bidirectional ACO intelligent fire evacuation route optimization[J]. *Journal of Ambient Intelligence and Smart Environments*, 2022, **14**(4): 1 − 19. DOI: 10.3233/AIS − 220620.

[13] Xia W, Cao K, Hu Q C. Ancient village fire escape path planning based on improved ant colony algorithm[C]// *IOP Conference Series*: *Earth and Environmental Science*. Chengdu, China, 2017, **69**(1): 012074. DOI: 10.1088/1755-1315/69/1/012074.

[14] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// *IEEE/CVF International Conference on Computer Vision* (*ICCV*). Montreal, Canada, 2021: 10012 − 10022. DOI: 10.1109/ICCV48922.2021.00986.

[15] Luo M, Xu L, Yang Y, et al. Laboratory flame smoke detection based on an improved YOLOX algorithm[J]. *Applied Sciences*, 2022, **12**(24): 12876. DOI: 10.3390/app122412876.

[16] Liu S, Qi L, Qin H, et al. Path aggregation network for

instance segmentation[C]//*IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Salt Lake City, UT, USA, 2018: 8759 – 8768. DOI: 10.1109/CVPR.2018.00913.

[17] Ali M R, Ma W X. New exact solutions of Bratu Gelfand model in two dimensions using Lie symmetry analysis[J]. *Chinese Journal of Physics*, 2020, **65**: 198 – 206. DOI: 10.1016/j.cjph.2020.01.008.

[18] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]// *Proceedings of the European Conference on Computer Vision* (*ECCV*). Munich, Germany, 2018: 3 – 19. DOI: 10.1007/978-3-030-01234-2_1.

[19] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]// *IEEE Conference on Computer Vision and Pattern Recognition*(*CVPR*). Honolulu, HI, USA, 2017: 936 – 944. DOI: 10.1109/CVPR.2017.106.

[20] Karur K, Sharma N, Dharmatti C, et al. A survey of path planning algorithms for mobile robots[J]. *Vehicles*, 2021, **3**(3): 448 – 468. DOI: 10.3390/vehicles3030027.

[21] Wu X, Lei Y, Tong X, et al. A Non-rigid hierarchical discrete grid structure and its application to UAVs conflict detection and path planning[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2022, **58**(6): 5393 – 5411. DOI: 10.1109/TAES.2022.3170323.

基于 YOLOX-Swin 的高效建筑火灾烟雾检测和疏散模拟方法

徐　照　戴天琦

（东南大学土木工程学院，南京 211189）

摘要：为了实现高效的建筑火灾应急救援疏散，分析了将目标检测技术应用于建筑火灾应急处置的可能性. 将目标检测算法应用于火灾预警阶段，将 Transformer、卷积神经网络 CNN 和轻量级注意力机制模块 CBAM 相结合，对火焰和烟雾局部和全局特征进行提取，提高目标检测算法的精度并实现对火灾发生位置的快速定位. 提出一种用于路径搜索的改进的蚁群算法，对启发函数和信息素挥发系数进行改进. 在案例中，建立栅格图模型，结合定位信息，通过仿真模拟的方式验证方法的有效性. 结果表明：相比与 YOLOX 算法，YOLOX-Swin 模型平均精度提高 1.5%；改进蚁群算法降低了传统蚁群算法的搜索范围，提高模型的收敛速度，有效避免了模型陷入局部最优解的困境. 将火灾预警和火灾人员疏散相结合，建立完整的建筑火灾应急处置方案.

关键词：计算机视觉；自注意力；蚁群算法；火灾动力学模拟

中图分类号：TU998.1；TP391.9