

---

## Footprints Walked: A Dialogue Between the YOLO and the Stairs

### Summary

Carved stones are often used as building materials, symbolizing stability and eternity, and carrying the weight of history and cultural imprints. This paper establishes the **FSM-YOLO wear identification model** and the **staircase information judgment model based on the YOLO algorithm** by improving and training the YOLO model, aiming to help archaeologists infer key information such as the historical background, usage frequency, and construction time of buildings.

First, this paper collected a large number of actual staircase photos and preprocessed and augmented them for subsequent model training.

For Task 1, we improved the existing YOLOv8n model to establish the **FSM-YOLO wear identification model**. First, we added the FPN structure on the basis of YOLOv8n, enabling the model to better recognize the degree of step wear. Then, inspired by the multi-branch structure, we introduced the **Squeeze-and-Excitation Attention (SA)** mechanism to dynamically adjust the attention weights, thereby better identifying the features of the staircase wear area. Next, we introduced a distance-weighted mechanism to the Intersection over Union (IoU), defining the **Mean Distance-Weighted IoU Loss (MDP-IOU)** loss function, which enhanced the model's fine matching ability for staircase wear information. Finally, we annotated the collected images, using 80% as the training set and 20% as the test set to train the FSM-YOLO wear identification model. The model training and testing effects are shown in **Figure 12**, and we answered the three questions in Task 1 based on the identified features.

For Task 2, we established a **staircase information judgment model based on the YOLO algorithm**. First, according to Assumption 2, we collected the measurement data that we considered important. Then, we conducted a mechanism analysis of the step wear situation and combined it with a simulation of the wear coefficient  $k$  to construct a staircase wear model. Next, we combined the staircase wear model with the **FSM-YOLO wear identification model** to obtain the **staircase information judgment model based on the YOLO algorithm**. After that, we annotated the collected image data, using 80% as the training set and 20% as the test set to train the model, and proposed using the **C-14 dating method** to determine the source of the staircase materials. Finally, we presented the model training and testing effects as shown in **Figure 14**, and answered the five questions in Task 2 based on the identified information.

Finally, this paper conducted a **sensitivity and robustness analysis** on the model and generalized it, verifying the model's stability and anti-interference capability, as well as its universality in multiple scenarios.

**Keywords:** YOLO algorithm; FSM-YOLO; SA mechanism; MDP-IOU; C-14 dating method

## Contents

1 Introduction .....	3
1.1 Problem Background .....	3
1.2 Restatement of the Problem .....	3
1.3 Literature Review .....	4
1.4 Our Work .....	4
2 Assumptions and Justifications .....	5
3 Notations .....	6
4 Data Preprocessing and Enhancement .....	6
4.1 Data Preprocessing .....	6
4.1.1 Image Preprocessing .....	6
4.1.2 Data Annotation .....	7
4.2 Data Enhancements .....	7
4.3 Image Normalization .....	8
5 FSM-YOLO Wear Recognition Model .....	8
5.1 The establishment of the FSM-YOLO wear recognition model .....	9
5.1.1 FPN structure .....	9
5.1.2 Squeeze-and-Excitation Attention (SA) Mechanism .....	10
5.1.3 MDP-IOU Loss Function .....	11
5.2 Training and Testing of the FSM-YOLO Wear Recognition Model .....	13
5.2.1 Usage Frequency .....	14
5.2.2 Walking Direction .....	14
5.2.3 Number of Users .....	15
5.2.4 Image Annotation .....	15
5.3 Results of the FSM-YOLO Wear Recognition Model .....	15
6 Staircase Information Judgment Model Based on YOLO Algorithm (SIBM-YOLO) .....	17
6.1 Staircase Wear Model .....	18
6.1.1 Wear rate coefficient $k$ .....	18
6.2 Model training .....	19
6.2.1 Wear amount .....	19
6.2.2 Specific usage frequency .....	19
6.2.3 Maintenance conditions .....	19
6.2.4 Image annotation .....	19
6.3 Material origin .....	20
6.4 Results display .....	20
7 Sensitivity and Robustness Analysis .....	22
7.1 Sensitivity Analysis .....	22
7.2 Robustness Analysis .....	23
8 Model Evaluation and Further Discussion .....	23
8.1 Model Evaluation .....	23
8.2 Further Discussion .....	24
References .....	24
Report on Use of AI .....	26

# 1 Introduction

## 1.1 Problem Background

Stone, known for its strength and durability, has been widely used in construction since ancient times, especially in historic temples and churches. Over time, however, stone is subject to wear due to prolonged use and environmental influences, with wear on staircases being particularly pronounced. The central areas of stair treads are often more worn than the edges, sometimes showing depressions or curvatures. These wear patterns not only record traces of human activity but also provide important clues for the historical evolution of architecture. The wear patterns on staircases reflect building usage frequency,

movement directions, and behavioral patterns, thus holding significant research value.



Figure 1: Worn Staircase

However, extracting this information presents various challenges, such as the lack of historical records, the impact of material differences on wear analysis, and the obfuscation of original features due to renovation activities. Therefore, there is an urgent need to develop scientifically efficient analytical tools, using non-destructive and cost-effective mathematical models to interpret wear characteristics. This study aims to unveil the usage history of ancient buildings through such modeling frameworks and advance the scientific interpretation of human behavioral patterns, offering new perspectives and methodologies for archaeological research.

## 1.2 Restatement of the Problem

Given the background information and constraints provided in the problem statement, our task is to develop a mathematical model that can be generalized to various scenarios of worn stairs. The model should address the following objectives:

Task 1: Determine the usage patterns of stairs:

- Estimate the frequency of use over time.
- Determine whether users tend to walk in a particular direction (e.g., going up or down).
- Analyze whether people tend to walk in a single file or side by side.
- Analyze whether the stairs are subject to high-intensity use over short periods of time or low-intensity use over long periods of time.

Task 2: Investigate the construction history of the stairs and human activity patterns over time:

- Estimate the approximate construction date of the stairs and assess the reliability of that estimate.
- Detect signs of repair, renovation, or modification of the stairs over time.
- Evaluate whether the wear pattern is consistent with the assumed material source (e.g., stone from a specific quarry or wood from a specific tree species).

### 1.3 Literature Review

In recent years, research based on step wear has been widely applied to the archaeological study of historical buildings and the analysis of human behavioral patterns. By quantitatively modeling wear patterns, researchers can uncover insights into the frequency of building use, directional movement preferences, and group behavior habits. This study proposes a novel approach leveraging machine learning techniques to develop the FSM-YOLO model, based on an improved YOLOv8<sup>[1]</sup>, to achieve automatic detection and classification of step wear regions<sup>[2]</sup>. Additionally, by integrating the FPN (Feature Pyramid Network) architecture and SA (Squeeze-and-Excitation Attention) mechanism<sup>[3]</sup>, the model significantly enhances multi-scale feature extraction capabilities, improving the accuracy of wear degree and walking direction recognition.

In the application of machine learning methods, the study employs data augmentation techniques to expand the diversity of staircase image datasets and uses an improved MDP-IOU<sup>[2]</sup> loss function to optimize bounding box prediction accuracy. Through the lightweight design of YOLOv8n, the model effectively improves detection efficiency while reducing resource consumption. Furthermore, this research combines biomechanical modeling and finite element analysis to propose a step lifespan prediction model based on wear rate coefficients. This model provides archaeologists with critical inferences regarding material sources, usage history, and renovation activities.

This study demonstrates the deep integration of machine learning with traditional archaeological research, offering a scientific tool to explore the construction and usage patterns of historical buildings.

### 1.4 Our Work

On the basis of the above analysis, we have carried out our work. Our working framework is shown in Figure 2.

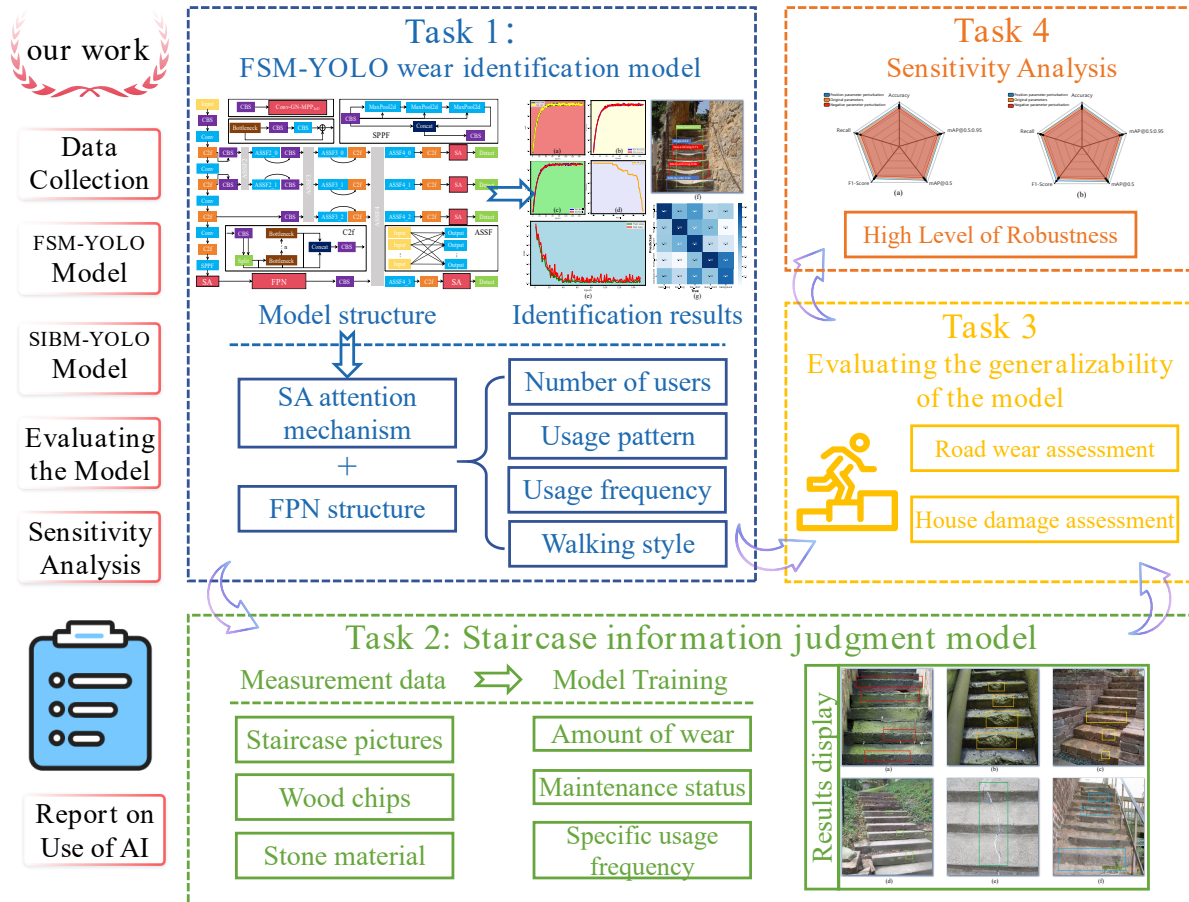


Figure 2: Our Work Overview Schematic Diagram

## 2 Assumptions and Justifications

Considering that practical problems always contain many complex factors, first of all, we need to make reasonable assumptions to simplify the model, and each hypothesis is closely followed by its corresponding explanation:

▼ Assumption 1: Assuming the YOLOv8n model can recognize the degree of wear.

▲ Explanations: While recognizing the wear of the steps, the YOLOv8n model will mark out the areas of different wear levels. By analyzing the size, depth and other characteristics of these areas, the usage frequency of the area can be inferred.

▼ Assumption 2: Assuming archaeologists have access to the relevant structures.

▲ Explanations: Archaeologists can directly access and measure the buildings, ensuring that we can obtain reliable and meaningful data to train the model.

▼ Assumption 3: Assuming that the frequency of wear is related to the usage pattern.

▲ Explanations: If certain areas of the steps show more concentrated wear, it may indicate that the staircase has experienced a higher frequency of use in a short period of time, whereas if the wear is more uniform, it suggests that the use has been more dispersed.

### 3 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description	Unit
$s$	Number of channel statistics generated by pooling	—
$X_g$	The $g^{\text{th}}$ group element of the division	—
$F_{gp}(X_{ki})$	Global average pooling function	—
$F_c$	Full Connection Layer Functions	—
$X'_k$	Number of recovery channels	—
$k$	coefficient of wear	—
$W$	Amount of wear and tear on stairs	$cm^2$
$f$	Frequency of use per unit of time	times/day
$N(t)$	The amount of C-14 remaining after time $t$	—
$N_0$	The amount of C-14 in modern wood	—
$T_{1/2}$	Half-life	year

**Note:** There are some variables that are not listed here and will be discussed in detail in each section.

## 4 Data Preprocessing and Enhancement

### 4.1 Data Preprocessing

We collected 201 pictures of staircases originating from actual photographs taken. To ensure the diversity and representativeness of the data, the collected pictures of staircases cover different staircase designs, environmental conditions, and use by people at different times of the day. Each image contains characteristics such as the wear area of the staircase, the activity of the user, the material of the staircase, etc.

#### 4.1.1 Image Preprocessing

Before data enhancement, the image needs to be uniformly formatted and standardized so that the model can effectively learn the features in the image. Specific preprocessing steps include:

- **Image format unification:** All images are formatted in JPG format to ensure consistency in the training process.
- **Image resizing:** As the YOLO model demands that the input image to have a fixed size, the image is resized to a uniform 224x224 pixels, ensuring that sufficient image detail is

retained.

- **Image centering:** To avoid the distortion of the image, the original proportion is maintained while the filling method is used to ensure that the main content of the image is located in the center of the area.

#### 4.1.2 Data Annotation

In this study, we manually labeled it with information including the frequency of stair usage for every image, the direction of travel, the number of people using the staircase, and the material.

### 4.2 Data Enhancements

In machine learning tasks, the problems of data imbalance and limited samples can be mitigated by data augmentation techniques. Data augmentation refers to performing various transformations (e.g., rotating, flipping, scaling, cropping, etc.) on the existing data to generate new training samples, thus increasing data diversity and improving the generalization ability of the model.

We use the ImageDataGenerator class from the TensorFlow library in Python to implement data enhancement. Here are the specific parameters for data enhancement:

- **Rotation range:** Maximum rotation angle is 20 degrees to simulate staircase images at different shooting angles.
- **Shear range:** The maximum shear range is 20 degrees, which increases the shape diversity of images.
- **Zoom range:** the maximum zoom range is 20%, simulating staircase images with different distances and scales.
- **Horizontal Flip:** Flip the image horizontally with a 50% probability to increase the symmetry of the image.
- **Generate samples:** 10 sample images are generated for each image.

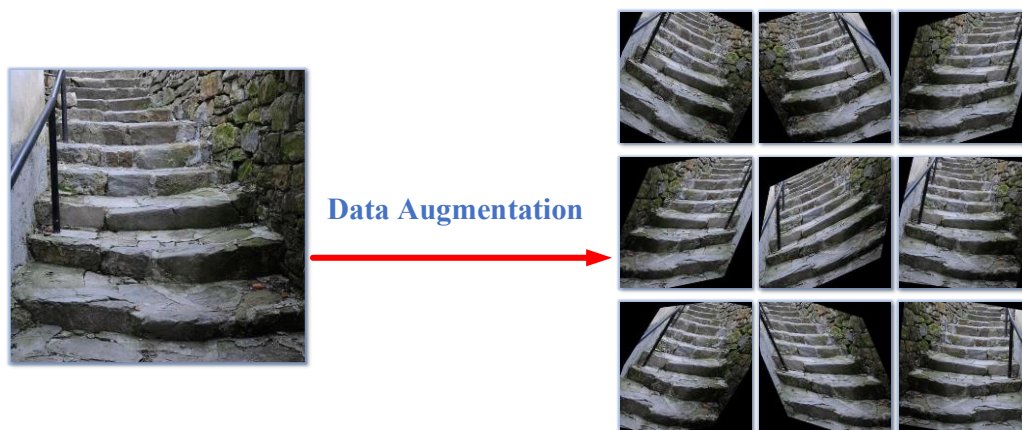


Figure 3: Schematic diagram of data enhancement

The parameter settings introduce randomness in the data enhancement process and increase the diversity of the training data to ensure that the subsequent model is better adapted to road images in different scenarios, thus improving the model's ability to recognize staircase situations.

### 4.3 Image Normalization

Before the image is input to the model, we normalize the image by scaling the pixel values of the image by adjusting the rescale parameter with the following equation:

$$x' = \frac{x}{255} \quad (1)$$

Where  $x$  denotes the original pixel value and  $x'$  denotes the normalized pixel value. The normalization process transforms the pixel values of the image from the range of 0-255 to the range of 0-1. This normalization significantly enhances the stability and convergence speed during model training.

## 5 FSM-YOLO Wear Recognition Model

YOLOv8 is a target detection algorithm released by Ultralytics in January 2023. Compared with other algorithms in the YOLO series, YOLOv8 focuses more on small target detection and has also optimized detection accuracy and speed. According to the scale of the model, YOLOv8 is divided into five versions: n, s, m, l, and x. For resource-constrained embedded devices, deploying YOLOv8 may consume a lot of resources. Therefore, this paper chooses YOLOv8n, which has a smaller number of parameters and computational amount, as the basic framework for research. The structure diagram of YOLOv8n is as follows:

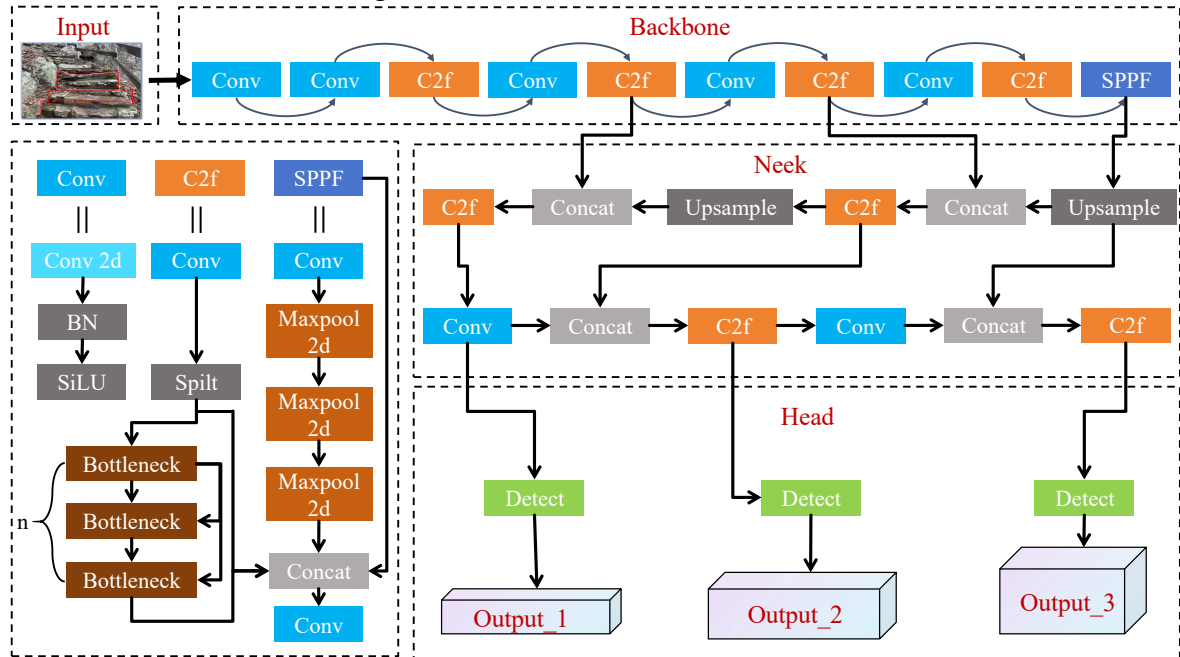


Figure 4: YOLOv8n Structure Diagram



The Backbone is responsible for feature extraction, using structures such as the C2f module to efficiently extract features; the Neck fuses features from different levels through the Path Aggregation Network (PAN) structure, enhancing the feature expression ability; the Head part is responsible for classification and bounding box regression, using techniques such as distribution focal loss and CIoU to improve detection accuracy. Overall, YOLOv8n achieves efficient feature extraction and accurate object detection while maintaining a lightweight structure.

## 5.1 The establishment of the FSM-YOLO wear recognition model

We have improved YOLOv8n to build the FSM-YOLO wear recognition model. First, we added the FPN structure on the basis of YOLOv8n to make the model better recognize the degree of step wear. Then we added the SA attention mechanism, so that the model can focus on recognizing the degree of step wear and discard external interference. Finally, we improved the loss function, defined the MDP-IoU loss function through distance weighting, further optimized the calculation of IoU, and enhanced the model's ability to finely match targets. The framework of the FSM-YOLO wear recognition model is shown in the figure below:

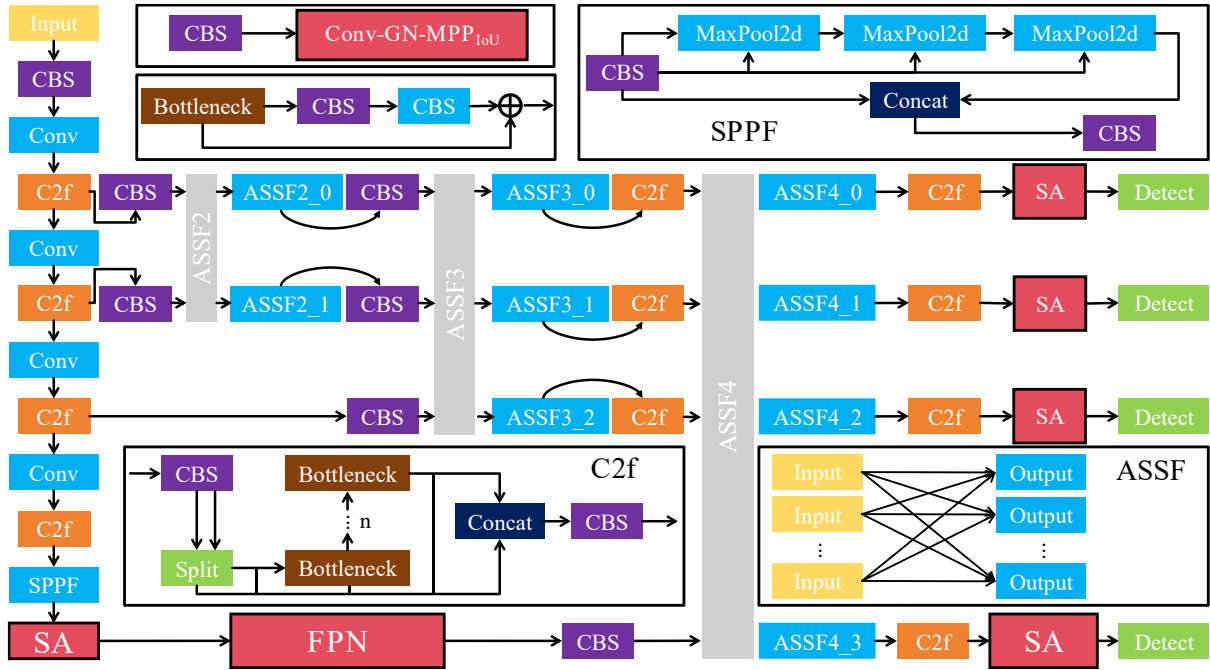


Figure 5: FSM-YOLO Wear Recognition Model Structure Diagram

### 5.1.1 FPN structure

The degree of step wear shows significant differences due to environmental and human factors. In step images, from low-level features such as edges, textures, and colors to more abstract high-level features such as target shapes and structures, multi-scale characteristics are shown. At the same time, the wear area is also different in size, with both small-scale local wear and large-scale severe wear. The FPN (Feature Pyramid Network) network structure, with its

unique top-down and lateral connection methods, can effectively extract and fuse feature information of different scales. This makes it perform well in detecting and recognizing step wear areas of different sizes, and can significantly enhance the precision and accuracy of detection. Common FPN structures include unidirectional FPN, PAFPN, BiFPN, etc. <sup>[4]</sup>, and the specific structures are shown in the figure below:

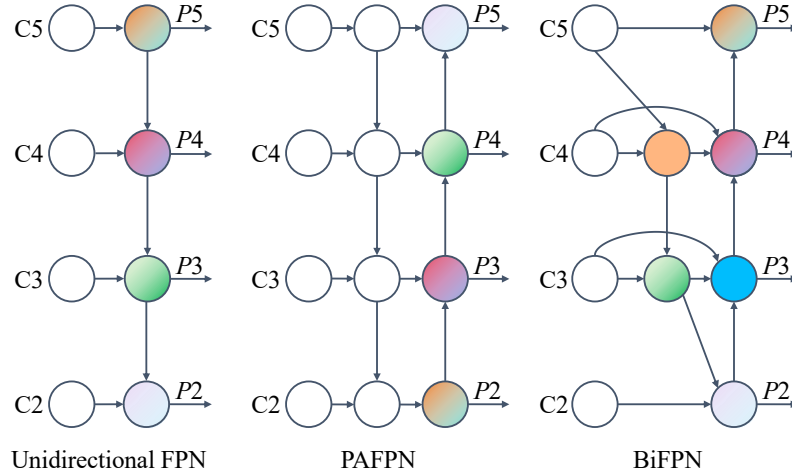


Figure 6: Schematic Diagram of Various FPN Structures

Among them, C1 - C5 and P1 - P5 represent the different levels of feature maps before and after FPN processing, respectively. The feature maps at each level of the FPN structure correspond to different levels of detail and overall information. The low-level feature maps contain rich detail information, which is very effective for detecting subtle wear on steps; the high-level feature maps have stronger semantic information, which can help identify the structure and location information of the steps, and thus judge the impact of wear on the overall structure.

### 5.1.2 Squeeze-and-Excitation Attention (SA) Mechanism

The SA mechanism is a module used to enhance the feature representation ability of Convolutional Neural Networks (CNNs). Its core idea is to introduce channel attention mechanism (SE)<sup>[5]</sup> to adaptively re-scale the weights of each feature channel, thereby enhancing the network's perception of key information. However, when processing image data, SE models cross-dimensional interaction to extract channel attention but ignores spatial features; Efficient Channel Attention (ECA)<sup>[6]</sup> achieves performance gains with minimal parameter addition, but it cannot well handle global context dependencies; Coordinate Attention (CA)<sup>[7]</sup> considers both channel and spatial dimensions, but it has a large computational overhead and cannot capture long-range dependencies; Efficient Multi-scale Attention (EMA)<sup>[8]</sup> can enhance the model's focus on important information, alleviate long-term dependency issues, but it has a high computational complexity and may cause overfitting problems. SA is an efficient attention fusion mechanism, inspired by the idea of multi-branch structure, it first groups the channel dimensions into mul-

multiple sub-features, and then applies spatial and channel attention mechanisms to each sub-feature simultaneously<sup>[2]</sup>. Finally, the Concat module is used to aggregate information within the group and achieve information communication between different sub-features, which helps the model better identify the wear of the steps, its structure is shown in the figure below:

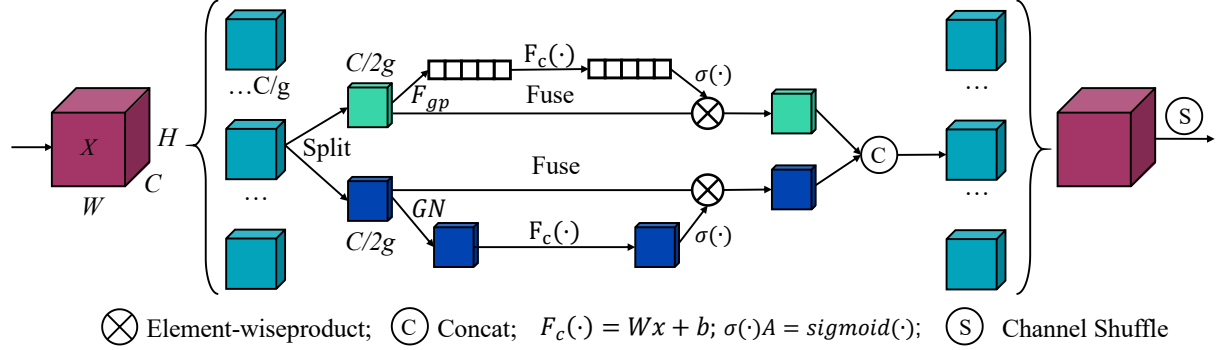


Figure 7: Structure Diagram of SA Attention Mechanism

When the feature  $X \in R^{C \times H \times W}$  ( $C$  is the number of channels,  $H$  is the height,  $W$  is the width) passes through the SA attention, it will first be divided into  $g$  groups  $X = [X_1, \dots, X_g]$  in the channel dimension, where  $X_k \in R^{\frac{C}{g} \times H \times W}$ , in each Unit, the initial  $X_k$  is Split into two branches  $X_{k1}$  and  $X_{k2} \in R^{\frac{C}{2g} \times H \times W}$  and spatial attention and channel attention are used respectively, and the groups are in parallel. For the channel attention branch, the channel statistics  $s$  are first generated through global average pooling, and its formula is:

$$s = F_{gp}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{k1}(i, j) \quad (2)$$

where  $F_{gp}(X_{k1})$  represents the global average pooling function, and then a feature is created through the sigmoid activation function for accurate adaptive selection, with the final output being:

$$X'_{k1} = \sigma(F_c(s)) \cdot X_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \quad (3)$$

where  $W_1, b_1 \in R^{\frac{C}{2g} \times 1 \times 1}$ , and  $F_c$  represent the fully connected layer functions,  $\sigma$  is the activation function sigmoid, and  $s$  is the result after global average pooling.

For the spatial attention branch, first apply Group Normalization to  $X_{k2}$  to obtain spatial statistics, and enhance it to get the final output:

$$X'_{k2} = \sigma(W_2 \cdot GN(X_{k2}) + b_2) \cdot X_{k2} \quad (4)$$

Where  $W_2, b_2 \in R^{\frac{C}{2g} \times 1 \times 1}$ ,  $GN$  is Group Normalization.

Then use the Concat module to concatenate the results of the two branches to restore the number of channels  $X'_k = [X'_{k1}, X'_{k2}] \in R^{\frac{C}{g} \times H \times W}$ , and finally use the channel shuffle operation to enhance the information interaction between channels, achieving cross-group information exchange along the channel dimension.

### 5.1.3 MDP-IOU Loss Function

The MDP-IOU Loss Function (Mean Distance-Weighted IoU Loss) is a loss function used

for object detection, aimed at improving the performance of the Intersection over Union (IoU) loss, especially when dealing with targets of different sizes or shapes. It further optimizes the calculation of IoU by introducing a distance-weighting mechanism, enhancing the model's ability to finely match the wear information of the stairs.

### ● IoU Loss Function

The IoU loss function is a standard measure of the overlap between two predicted boxes and ground-truth boxes. The definition of IoU is shown in the figure below:

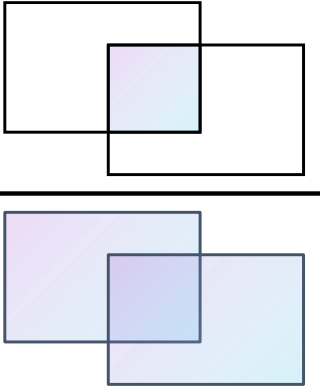
$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$


Figure 8: Definition of IoU Loss Function

Usually, the training objective of object detection models is to maximize the IoU, so as to make the predicted box as close to the ground-truth box as possible.

### ● The working principle of MDP-IoU

The MDP-IoU loss function is an improvement on the traditional IoU loss, with the core being distance weighting. The purpose of this method is to make the loss of the predicted box farther from the ground-truth box larger, and the loss of the predicted box closer to the ground-truth box smaller, thereby increasing the model's requirement for the precision of the bounding box, as shown in the figure below:

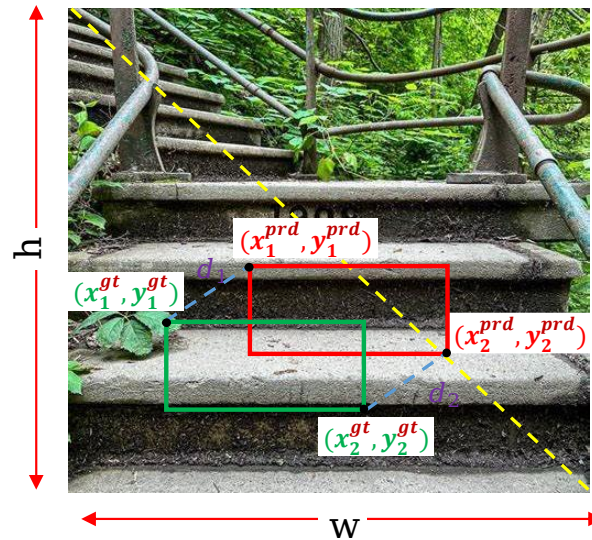


Figure 9: Working Principle of MDP-IoU

where  $A_{\text{prd}}((x_1^{\text{prd}}, y_1^{\text{prd}}), (x_2^{\text{prd}}, y_2^{\text{prd}}))$  represents the predicted bounding box  $A$ ,  $B_{\text{gt}}((x_1^{\text{gt}}, y_1^{\text{gt}}), (x_2^{\text{gt}}, y_2^{\text{gt}}))$  represents the ground-truth bounding box  $B$ ,  $w$  and  $h$  represent the width and height of the image, and  $d_1$  and  $d_2$  are intermediate parameters.

During the training phase, the predicted bounding box will approach the ground-truth bounding box by minimizing the loss function, and the formula for the loss function is:

$$Loss = \min_{\Theta} \sum_{B_{\text{gt}}} L(B_{\text{gt}}, B_{\text{prd}} | \Theta) \quad (5)$$

where  $\Theta$  is the given model parameter, and  $L(B_{\text{gt}}, B_{\text{prd}} | \Theta)$  is the loss calculated for each ground-truth bounding box  $B_{\text{gt}}$  and predicted bounding box  $B_{\text{prd}}$ . In this process, the expressions for the intermediate parameters are first obtained based on the geometric relationships of the coordinates of the two boxes:

$$d_1^2 = (x_1^{\text{prd}} - x_1^{\text{gt}})^2 + (y_1^{\text{prd}} - y_1^{\text{gt}})^2 \quad (6)$$

$$d_2^2 = (x_2^{\text{prd}} - x_2^{\text{gt}})^2 + (y_2^{\text{prd}} - y_2^{\text{gt}})^2 \quad (7)$$

Then, based on the definition of the IoU loss function, the expressions for  $MDP_{\text{IoU}}$  and  $Loss_{MDP_{\text{IoU}}}$  are obtained as:

$$MPD_{\text{IoU}} = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (8)$$

$$Loss_{MPD_{\text{IoU}}} = 1 - MPD_{\text{IoU}} \quad (9)$$

The FSM-YOLO model established in this paper uses  $MPD_{\text{IoU}}$  as the loss function to replace the original  $C_{\text{IoU}}$  of the YOLOv8n model, incorporating factors such as non-overlapping area, center point distance, and width and height deviations into the loss calculation factors. Without increasing the computational load, the model's detection performance is further improved.

## 5.2 Training and Testing of the FSM-YOLO Wear Recognition Model

We used annotated staircase images as the training dataset, with 80% of the dataset as the training set and 20% as the test set to train and test the FSM-YOLO wear recognition model. To accurately evaluate the performance of the improved model, while considering the need for lightweight, this paper mainly uses Precision (P), Recall (R), and Mean Average Precision (mAP) as the evaluation indicators of the model.

The operating system used for training the model is Windows 10, the GPU is NVIDIA RTX 2060 Super, the development platform is PyCharm, the compilation language is Python 3.13.1, the deep learning framework is PyTorch, and the GPU acceleration library is CUDA 12.0.

The image size for training the model is  $224 \text{ pixels} \times 224 \text{ pixels}$ , the learning rate is 0.01, the batch size is 16, the momentum is set to 0.937, the decay coefficient is 0.0005, and the number of training epochs is 150. The pseudocode for model training is as follows:

Table 2: Pseudocode for FSM-YOLO Model Training

<b>Algorithm 1 FSM-YOLO model training</b>	
<b>Input:</b>	epochs , imgsz , device , batch, cache , lr0
<b>Output:</b>	Accuracy ,Recall , Mean Average Precision
1	<b>Begin</b>
2	Initialize the FSM-YOLO model
3	Define the MDP-IoU Loss Function
4	Define the optimizer
5	<b>for</b> epoch <b>in range</b> (Training epochs) <b>do</b>
6	Set training mode
7	<b>for</b> Batch images, batch image labels <b>in</b> Test dataset <b>do</b>
8	Adjust image size
9	Predict target information
10	Calculate loss value
11	Zero out gradients
12	Perform backpropagation
13	Update parameters
14	<b>end for</b>
15	<b>end for</b>
16	Output training results
17	<b>End</b>

### 5.2.1 Usage Frequency

We use the degree of wear on the staircase to describe its usage frequency. Based on the varying degrees of wear, the usage frequency of the staircase is divided into four scenarios: short-term use by a small number of people (few and short), long-term use by a large number of people (many and long), short-term use by a large number of people (many and short), and long-term use by a small number of people (few and long).

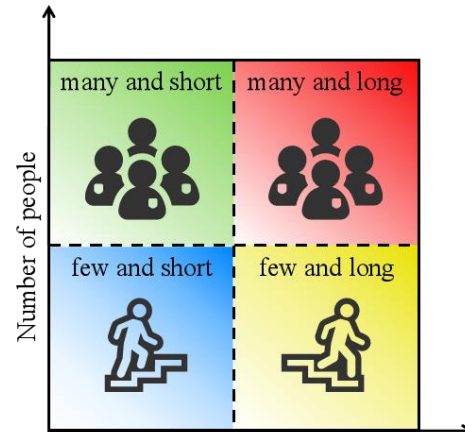


Figure 10: Four Usage Frequencies

Among them, the wear degree is the smallest for a small number of people using it in the short term, and the wear degree is the largest for a large number of people using it in the long term. For a large number of people using it in the short term, the wear degree is concentrated, the depth is large, but the range is limited. For a small number of people using it in the long term, the wear degree is uniform, the depth is shallow, but the range is extensive.

### 5.2.2 Walking Direction

In our study, walking direction is defined into two categories: upward and downward. By



analyzing the wear of the stairs, we can further study the patterns of crowd movement, walking speed, and whether there is a clear preferred direction.

### 5.2.3 Number of Users

In this study, the number of people using the stairs is divided into two categories: walking side by side in pairs (two people walking side by side together) and walking in a single file (one person walking the stairs alone).

### 5.2.4 Image Annotation

We will label the images of usage frequency, walking direction, and number of users according to the above characteristics as the training set to train the FSM-YOLO wear recognition model, and the labeling results are shown in the figure below:



Figure 11: (a) Labeling indicates the wear of stairs used by a few people over a long period. (b) Labeling indicates the wear of stairs used by many people over a short period. (c) Labeling indicates the wear condition of the stairs in the upward direction. (d) Labeling indicates the wear condition of the stairs in the downward direction. (e) Labeling indicates the wear condition of the stairs when walking in a single file. (f) Labeling indicates the wear condition of the stairs when walking side by side in pairs.

## 5.3 Results of the FSM-YOLO Wear Recognition Model

We use Precision, Recall, Precision-Recall curve (PR curve), and Mean Average Precision (mAP) as the evaluation metrics for model training. The expression for Precision is as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

where TP is the number of samples predicted as positive and actually positive, and FP is the number of samples predicted as positive but actually negative. The expression for Recall is as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

where FN is the number of samples that are actually positive but not correctly predicted. The PR curve is a graph showing the relationship between Precision and Recall at different confidence thresholds. Mean Average Precision is the average of the Average Precision for all categories, a key metric for measuring the overall detection performance of the model. The higher the value, the better the object detection performance of the model. The formulas for AP and mAP are as follows:

$$AP = \int_0^1 \text{Precision}(R) dR \quad (12)$$

$$mAP = \frac{1}{N} \sum_N AP_i \quad (13)$$

where  $R$  represents the value of Recall,  $N$  is the total number of categories, and  $AP_i$  is the Average Precision of the  $i_{th}$  category.

We selected a representative staircase image from the test set to showcase the results of the FSM-YOLO wear recognition model, as shown in the figure below:

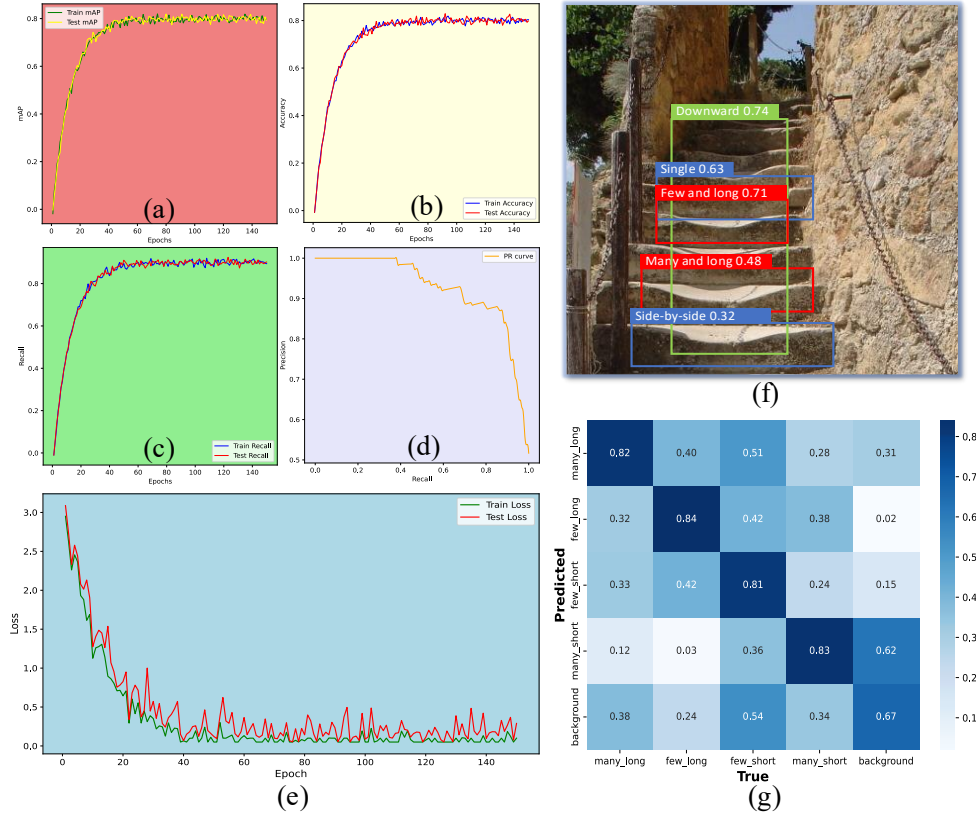


Figure 12: (a) is the mAP curve during model training. (b) is the Accuracy curve during model



training. (c) is the Recall curve during model training. (d) is the PR curve during model training. (e) is the loss function curve during model training. (f) is the staircase image recognized by the trained model. (g) is the confusion matrix.

From Figures (a), (b), (c), and (e), it can be seen that as training progresses, the model's various indicators gradually improve, indicating good training results. From Figure (d), it can be seen that the model has a good balance of precision and recall on various types of samples, indicating good performance in recognizing the wear and tear of the stairs. From Figure (f), it can be seen that the model has identified different stair usage frequencies (red), walking methods (green), and number of users (blue) in the stair image, and has given confidence levels, further enhancing the credibility of the model. From Figure (g), it can be seen that the values on the main diagonal of the confusion matrix are all above 0.6, indicating that the model performs well in recognizing stair wear information.

Based on the information identified by the model, the three questions in Task One can be answered:

- **For question 1:** As shown by the red box in Figure (f), the model has identified the usage frequency category of the stairs and provided a confidence level.
- **For question 2:** As indicated by the green box in Figure (f), the model has determined that the stairs are for descending, with a confidence level of 0.74, indicating a relatively high degree of reliability in the result.
- **For question 3:** As denoted by the blue box in Figure (f), the model has identified the number of people using the stairs, with a higher confidence level for single-file walking.

## 6 Staircase Information Judgment Model Based on YOLO Algorithm (SIBM-YOLO)

Based on Hypothesis 2, we have collected the measurement data that we consider important:

- Staircase images containing wear amount, usage frequency, and types of renovation, which are used to train the model established in Task 1.
- Residual C-14 in the wooden structures around the staircase, collected because archaeologists have access to the relevant buildings, to accurately calculate the age of the staircase.
- Wear coefficient  $k$  of various types of stone from surrounding quarries, collected to determine the source of the staircase materials.

### 6.1 Staircase Wear Model

In this study, we plan to estimate the usage years of the staircase by measuring the degree of wear on the staircase. The degree of wear is usually closely related to the usage frequency of the staircase, the characteristics of the material, and external environmental factors. Assuming that the wear amount is a function of time and usage frequency, the higher the usage frequency, the greater the wear rate of the staircase, which in turn affects the estimation of the usage years. The wear model we propose is as follows:

$$W = k \cdot f \cdot t \quad (14)$$

where  $W$  represents the wear amount of the staircase,  $k$  is the wear rate coefficient determined by the physical properties of the staircase material and surface hardness,  $f$  denotes the usage frequency per unit time, and  $t$  represents the usage time of the staircase. The usage time of the staircase can be calculated as follows:

$$t = \frac{W_{\text{measured}}}{k \cdot f_{\text{measured}}} \quad (15)$$

where the subscript measured represents the measured value.

### 6.1.1 Wear rate coefficient $k$

To accurately calculate the wear rate of the staircase material, this study employs the Finite Element Analysis (FEA) method for simulation. FEA can precisely simulate the deformation behavior of materials under friction, calculate the stress distribution in the contact area and the wear condition of the material, thereby providing a theoretical basis for the material's wear rate.

## ● Simulation

First, we used SolidWorks software to create the geometric models of the staircase materials (wood and stone). Then, we set appropriate material properties for the stone, including elastic modulus, Poisson's ratio, density, and hardness. The hardness of stones such as marble and granite is relatively high, and the differences in material properties will directly affect the calculation results of wear. We then used the Archard wear model to define the wear behavior of the material, and the basic formula of the Archard model is:

$$V = k \cdot \frac{F \cdot S}{H} \quad (16)$$

where  $V$  is the wear volume,  $F$  is the gravitational force of the human body on the step,  $S$  is the relative sliding distance, and  $H$  is the hardness of the material. Finally, the geometric model is meshed, and after substituting the numerical values, the wear rate coefficient  $k$  for marble is obtained as:  $8.3 \times 10^{-5}$ .

## 6.2 Model training

We used the annotated staircase images as the training dataset, with 80% of the dataset as

the training set and 20% as the test set to train the FSM-YOLO wear recognition model. The evaluation metrics, parameter settings, and environment configuration of the model are the same as in Task 1.

### 6.2.1 Wear amount

By annotating the worn areas in the staircase photos as the training set, the staircase photos contain information such as shooting distance and angle. The trained model can identify the wear amount based on the number of pixel points framed.

### 6.2.2 Specific usage frequency

Similar to the recognition of wear amount, we identify the usage frequency based on the size of the wear amount. Since the staircase photos contain information such as shooting distance and angle, the trained model can identify the specific usage frequency through the size of the wear amount.

### 6.2.3 Maintenance conditions

We categorize the maintenance conditions into two types: filling and repair. Filling refers to filling the cracked parts of the steps with substances such as cement and crack repair agents. Repair means repairing the damaged, fallen off, or cracked parts of the steps. We annotate the pictures of these two types of maintenance.

### 6.2.4 Image annotation

We annotated the images of wear amount, specific usage frequency, and maintenance conditions according to the above characteristics, and used them as the training set to train the FSM-YOLO wear recognition model. The annotation results are shown in the figure below:

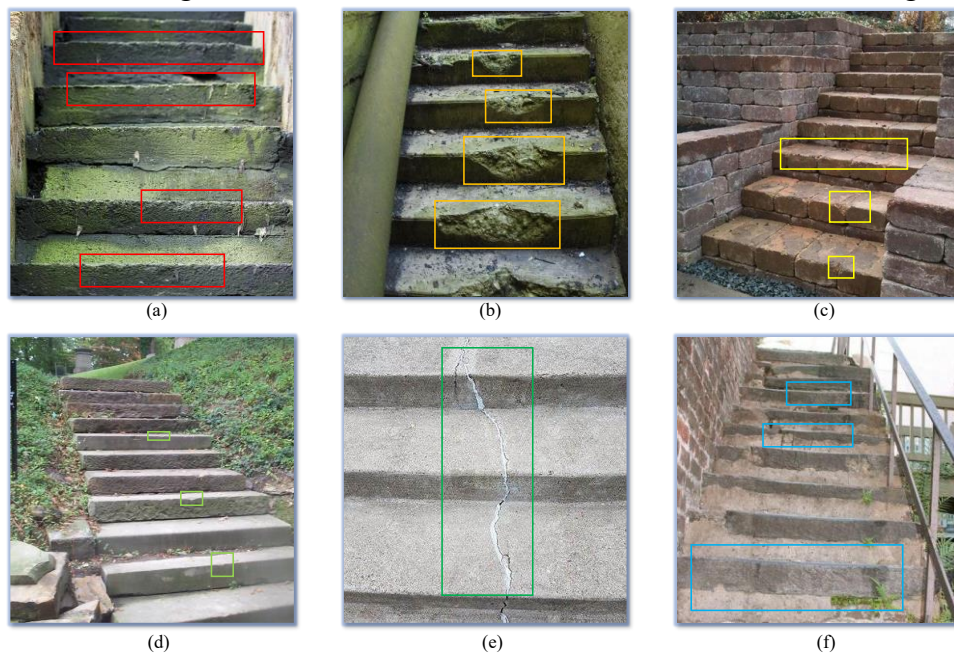


Figure 13: (a) The staircase with a wear amount of 5-10 cm<sup>2</sup> is annotated. (b) The staircase with

a wear amount of 20-30 cm<sup>2</sup> is annotated. (c) The staircase with a usage frequency of 30 times per day is annotated. (d) The staircase with a usage frequency of 10 times per day is annotated. (e) The staircase repaired by filling is annotated. (f) The staircase repaired by patching is annotated.

### 6.3 Material origin

Based on the carbon-14 content of the wood in the buildings surrounding the staircase that we have collected, we can accurately estimate the year of death of the wood, and thus infer the construction time of the wooden staircase. The decay of carbon-14 is an exponential decay process, following the formula:

$$N(t) = N_0 \cdot e^{-\lambda t} \quad (17)$$

where  $N(t)$  is the amount of C-14 remaining after time  $t$ ,  $N_0$  is the amount of C-14 in modern wood, and  $\lambda$  is the decay constant of C-14. The relationship between the half-life  $T_{1/2}$  of C-14 and the decay constant is:

$$T_{1/2} = \frac{\ln 2}{\lambda} \quad (18)$$

where the half-life of C-14 is approximately 5730 years, it can be concluded that:  $\lambda = 1.21 \times 10^{-4}$  per year. Based on the decay formula of C-14, taking the natural logarithm yields the age  $t$  of the wood as:

$$t = \frac{\ln \left( \frac{N_0}{N(t)} \right)}{\lambda} \quad (19)$$

Finally, by combining formulas (9) and (13), the wear coefficient  $k$  of the step material can be obtained. By comparing it with the wear coefficient of the materials from the attached raw material factory, the source of the staircase material can be determined.

### 6.4 Results display

Similar to the model training result presentation in Task 1, I choose Precision, Recall, Precision-Recall curve (PR curve), and Mean Average Precision (mAP) as the evaluation metrics for the model.

To visually demonstrate the training effectiveness of the FSM-YOLO wear recognition model, we selected a representative image of staircase wear from the test set, and the training results are shown in the figure below:

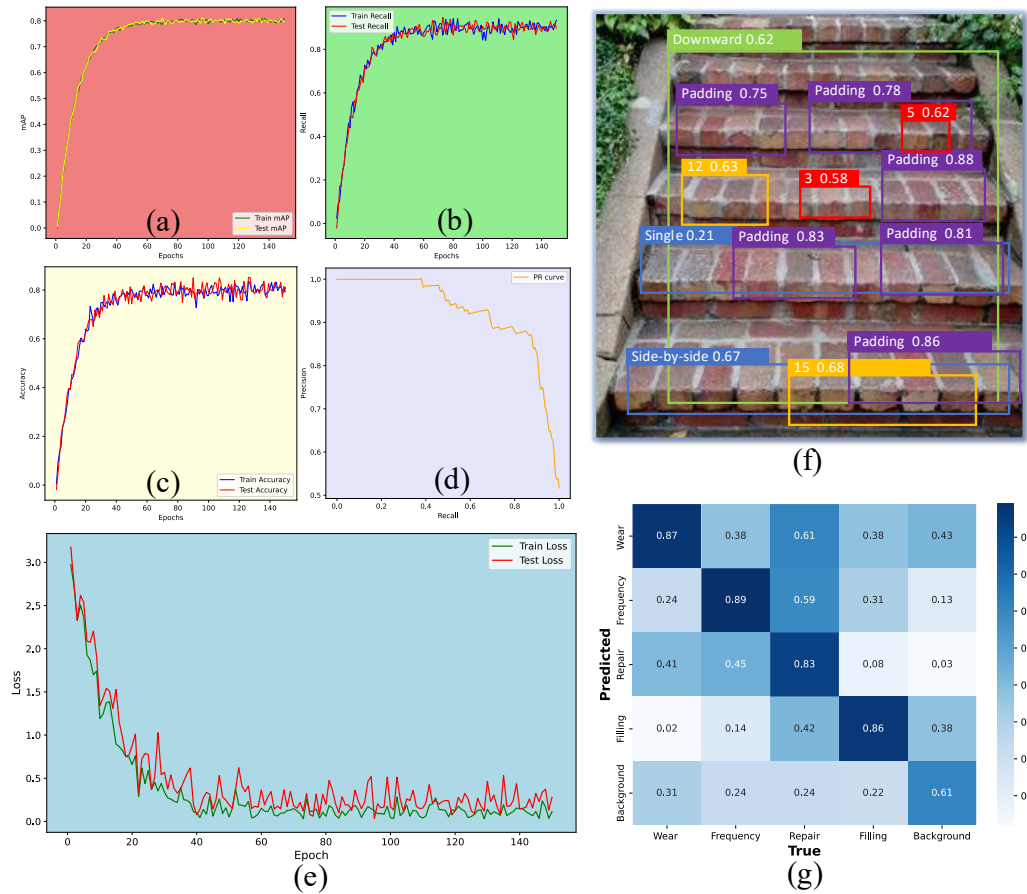


Figure 14: (a) is the mAP curve during model training. (b) is the Recall curve during model training. (c) is the Accuracy curve during model training. (d) is the PR curve during model training. (e) is the loss function curve during model training. (f) is the staircase image recognized by the trained model. (g) is the confusion matrix.

Similar to the training results in Task 1, it can be seen from the five figures on the left and the confusion matrix that as the training progresses, the various indicators of the model gradually improve, and the training effect is good. From Figure (f), it can be seen that the model recognizes different wear features of the staircase in the staircase image, such as wear amount (red), specific usage frequency (yellow), maintenance conditions (purple), walking direction (green), and number of users (blue), and provides confidence levels.

Based on the information identified by the model, the five questions in Task 2 can be answered:

- **For Question 1:** As shown by the red frame in Figure (f), the model identifies the wear amount of the staircase and provides a confidence level, which can effectively determine whether it is consistent with the available information.
- **For Question 2:** As shown by the red and yellow frames in Figure (f), the model identifies the usage frequency and wear amount. By substituting these into the staircase wear model in Section 6.1, the usage age can be determined. The reliability of the estimate can be

verified using the C-14 half-life method proposed in Section 6.3.

- **For Question 3:** As shown by the purple frame in Figure (f), the model can directly identify the maintenance conditions.
- **For Question 4:** By using the usage frequency and wear amount identified by the model, along with the age calculated in Question 2, the wear coefficient  $k$  can be calculated according to the staircase wear model. Comparing this with the wear coefficient of materials from surrounding quarries can determine the source of the material.
- **For Question 5:** The model trained in Task 1 can be utilized. Based on the definition of usage frequency in Task 1, the number of people using the staircase in a typical day can be directly obtained.

## 7 Sensitivity and Robustness Analysis

### 7.1 Sensitivity Analysis

The performance of deep learning models depends on the accurate values of their parameters. In practical deep learning application scenarios, model parameters may be affected by various factors and undergo minor changes. Since the weights of the model determine the degree of influence of input features on the output, and the bias is mainly used to adjust the output baseline of the model, to assess the sensitivity of the model to parameter changes, we adopt the method of adding random noise to the model parameters (weights and biases) for sensitivity analysis, with the noise following a Gaussian distribution. The training indicators of the model are shown in the figure below:

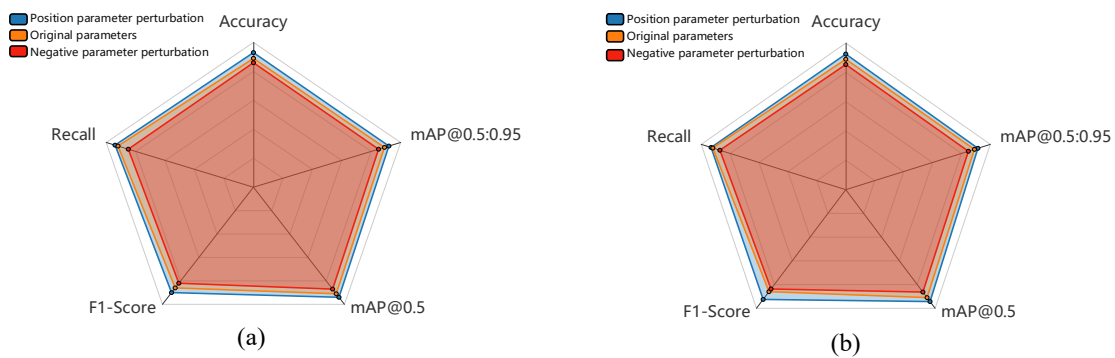


Figure 15: (a) is the change in various model performance evaluation indicators before and after adding disturbances to the model's weight parameters, (b) is the change in various model performance evaluation indicators before and after adding disturbances to the model's bias parameters.

From the above figure, it can be seen that after adding parameter disturbances to the mod-

el's weights and biases respectively, although the various indicators of the model have fluctuated accordingly, the magnitude of the fluctuations is within the acceptable range. Therefore, we believe that the model is not sensitive to external disturbances and has good stability.

## 7.2 Robustness Analysis

To test the generalization ability of the model, we consider conducting a robustness analysis on the FSM-YOLO model. Intentionally inject some malicious data into the test data, such as data containing incorrect annotations or abnormal targets. For example, in the training data set of the steps, add different numbers of unclear and obviously incorrectly annotated images, and then use the data set containing malicious data to train the model, observe the changes in the model recall rate and mean average precision (mAP) indicators. If the fluctuation of the indicators is small, it is considered that the model has strong robustness. The retraining results for the changed data set are shown in the figure below:

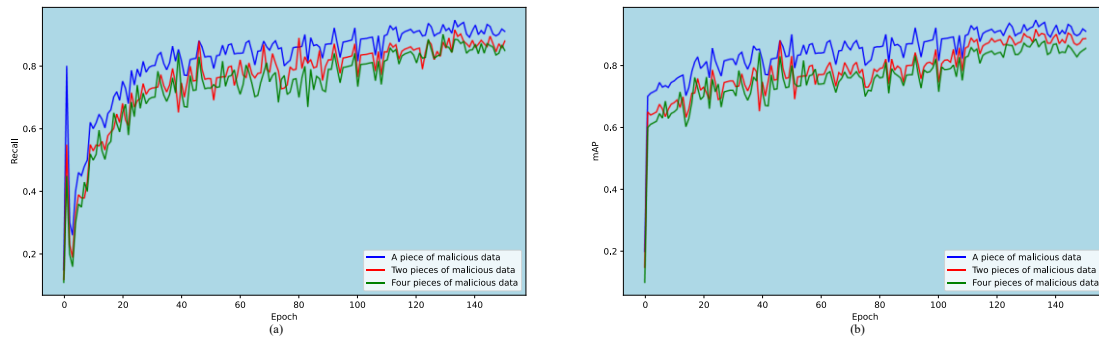


Figure 16: (a) is the recall rate of the model under different amounts of malicious data injection. (b) is the mean average precision (mAP) of the model under different amounts of malicious data injection.

From the above figure, we can conclude that after the introduction of malicious data, the fluctuations in the recall rate and mean average precision (mAP) of the model are relatively smooth, proving that the model has strong generalization ability and good robustness.

## 8 Model Evaluation and Further Discussion

### 8.1 Model Evaluation

- The FSM-YOLO wear recognition model and the SIBM-YOLO model for staircase information judgment based on the YOLO algorithm constructed in this study demonstrate multiple advantages. In terms of model performance, the FSM-YOLO model, by adding the FPN structure, introducing the SA attention mechanism, and improving the loss function, shows excellent performance in recognizing the degree of staircase wear, walking direction, and number of users. During the training process, the model's various indicators are

gradually optimized, with mAP, accuracy, and recall steadily increasing, and a good balance between precision and recall on different types of samples, with high values on the main diagonal of the confusion matrix, effectively identifying staircase wear information. The SIBM-YOLO model, by combining the staircase wear model, C-14 dating method, and material wear coefficient, accurately judges the service life, maintenance status, and material source of the staircase.

- From the perspective of methodological innovation, the integration of machine learning and traditional archaeological research provides a new method for the study of ancient architecture. Data augmentation techniques are used to expand the diversity of the dataset, enhancing the model's generalization ability; the improved MDP-IOU loss function optimizes the precision of bounding box prediction and reduces computational load. In practical application scenarios, the model performs well in processing staircase data under different scenarios, and sensitivity analysis shows that the model is not sensitive to parameter changes and has strong stability; robustness analysis shows that under malicious data interference, the model has small fluctuations in recall and mAP, and can be widely applied to various staircase research scenarios.
- The model still has certain limitations. In terms of data, although 201 staircase images were collected and enhanced, in complex scenarios, such as staircases in different architectural styles and extreme environmental conditions, the samples may not be rich enough, affecting the model's ability to recognize special cases. In addition, data annotation relies on manual work, which has subjectivity and annotation errors, and may reduce the accuracy of model training.

## 8.2 Further Discussion

Future research can be expanded in multiple directions. In terms of data collection and processing, the sample scope can be expanded to cover more types of staircase data, and automatic labeling technology or crowdsourced labeling can be used to reduce labeling errors. In terms of model optimization, more efficient feature extraction structures and attention mechanisms can be explored to reduce computational load and enhance recognition capabilities in complex scenarios; adaptive models can be developed to adjust parameters and structures according to different scenarios. The fusion of multiple models is also an important direction, combining deep learning models with traditional machine learning models to integrate the advantages of various models and improve model performance.

## References

- [1] Yi H, Liu B, Zhao B, et al. Small object detection algorithm based on improved YOLOv8 for remote sensing[J]. IEEE Journal of Selected Topics in Applied Earth Observations and



Remote Sensing, 2023.

- [2] Cheng Z, Gao L, Wang Y, et al. EC-YOLO: Effectual Detection Model for Steel Strip Surface Defects Based on YOLO-V5[J]. IEEE Access, 2024.
- [3] Bao Z. The UAV Target Detection Algorithm Based on Improved YOLO V8[C]//Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition. 2024: 264-269.
- [4] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.
- [5] Zhang H, Ma J. SDNet: A versatile squeeze-and-decomposition network for real-time image fusion[J]. International Journal of Computer Vision, 2021, 129(10): 2761-2785.
- [6] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.
- [7] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.
- [8] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

## Report on Use of AI

AI version: *OpenAI ChatGPT*(Sep 3, 2024 version, ChatGPT-4o)

Query 1: What are the wear characteristics of steps with different travel directions?

Output 1:

### Steps that are mainly used for walking upwards

- **Heavy wear on the front edge:** When people walk upwards, their feet will fall more on the front edge of the step, and their body's center of gravity and strength will be concentrated on this part, so the front edge of the step will be subjected to greater pressure and friction, and after long-term use, the front edge will show obvious wear and tear, and there may be deeper dents or wear marks.
- **Side wear is lighter:** since the walking direction is mainly upward, the sides are subject to relatively less friction and collision, so the side wear is usually not obvious.
- **Overall wear is distributed in a front-heavy and back-light manner:** from the viewpoint of the entire tread surface of the step, the degree of wear gradually decreases from the leading edge to the trailing edge, forming a front-deep and back-light wear gradient.

### Mainly downward-facing steps

- **Significant wear on the trailing edge:** When walking down the step, the foot will put more weight on the trailing edge of the steps. To maintain balance, the trailing edge will be subjected to more pressure and friction, resulting in faster wear on the trailing edge, which may result in dents or flaking of the material.
- **The leading edge may be slightly worn:** In the process of walking downwards, there may be some dragging or scraping of the leading edge when the foot leaves the step, so the leading edge will also have some slight wear, but usually less than the trailing edge.
- **Overall wear is rear-heavy and front-light:** wear on the tread surface is characterized by deep wear on the rear edge and shallow wear on the front edge.

### Steps that walk in both directions

- **Outstanding wear in the center:** Because of bi-directional walking, people's walking paths on steps are relatively concentrated in the center, so the center part of the steps will be subjected to more treading and friction.
- **Relatively light wear on both sides:** Compared with the center part, the two sides are used relatively less frequently, so the wear is relatively light, but with the increase in

use time, the two sides will also have a certain degree of wear.

- **Overall wear is deep in the center and shallow on both sides:** the wear pattern of the whole step tread shows a trend of larger depth of wear in the center and gradually becoming shallower on both sides.

AI version: *OpenAI ChatGPT*(Sep 3, 2024 version, ChatGPT-4o)

Query 2: What effect does the climatic environment around the steps have on the wear and tear of the steps?

Output 2:

### Light factor

- **Ultraviolet radiation:** long-term ultraviolet irradiation will make some step materials, will destroy the structure of the material, so that the molecular structure of the material changes, resulting in the material becoming brittle, and hard, losing the original elasticity and toughness, the surface is easy to crack and peel, thus accelerating the wear and tear of the steps.

### Humidity factor

- **Humid environment:** High humidity environment will make the step material damp, and reduce the strength and hardness of the material. At the same time, a humid environment also easily triggers microbial growth, microbial growth, and the metabolism process may have a corrosive effect on the surface of the steps, accelerating the wear and tear of the steps.
- **Rainfall:** Frequent rainfall will have a direct scouring effect on the surface of the steps, and the flow of rain will carry dust, sand, and other particles, which will cause friction and erosion on the surface of the steps under the drive of the water flow so that the materials on the surface of the steps will be worn out gradually.

### Biological factors

- **Plant growth:** If the environment around the steps is humid and there is suitable light and other conditions, there may be plants growing in the crevices or edges of the steps. The root system of plants will extend to the gap of the steps during the growth process. The growth and expansion of the root system will produce extrusion pressure on the steps, which will gradually enlarge the gap, resulting in the decline of the structural stability of the steps, and the surface materials will be easy to be loosened and detached, which will accelerate the wear and tear of the steps.

AI version: *OpenAI ChatGPT*(Sep 3, 2024 version, ChatGPT-4o)

Query 3: Is there some connection between the origin of the steps and age?

Output 3:

There is usually a relationship between step source and age, which is analyzed below in terms of different step source types:

### **Naturally occurring steps**

- Naturally occurring steps are often created gradually over a long period of geological time, for example, shaped by natural forces such as crustal movement, weathering of rocks, erosion by water, and so on. The age of these natural steps is generally very old, possibly millions or even hundreds of millions of years old, and their formation is closely related to the earth's geological evolution. For example, the natural steps in some mountainous areas formed by rock stratification and joints under long-term weathering and erosion, whose age corresponds to the period of the geological history of the area, have taken a long time to present the present form.

### **Artificially Constructed Steps**

- **Steps in Ancient Architecture:** In ancient times, steps were often used in palaces, temples, mausoleums, and other important buildings, and the construction of these steps was often related to a specific historical period and cultural background. Through historical documents, archaeological discoveries, etc. can be roughly deduced from their construction age, usually hundreds or even thousands of years of history. Due to the long construction time, and through the environmental changes and wear and tear of use in different historical periods, these steps carry a wealth of historical and cultural information, and their age is closely linked to the history of the ancient buildings to which they are attached.
- **Steps in Modern Architecture:** Steps constructed in the modern era have changed considerably in terms of materials and workmanship. These steps are relatively young, usually within a few decades. They are often built to meet the functional needs of modern buildings, such as residential, commercial buildings, public facilities, etc., the age of the steps is consistent with the time of completion of the building, and with time will gradually undergo a variety of impacts brought about by the natural environment and the process of use and change.