Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/133/2025.20799

Intelligent Monitoring and Safety Management System Construction of Manhole Cover Hidden Danger Based on YoSwin-Former Model

Jiajia Liu[1,a,*], Yu Zeng[1,b], Jiahao Wu[1,c]

[1]Finance and Economics University Of Jiangxi, Nanchang, China
a. 18050702759@163.com, b. 2671104571@qq.com, c. 3181530800@qq.com
**corresponding author*

Abstract: This paper focuses on the safety management of urban manhole covers. In view of the drawbacks of the traditional management model and the frequent occurrence of manhole cover accidents, an intelligent manhole cover hidden danger identification project is proposed. The project is dedicated to solving the problems of high cost, difficult hidden danger detection, and low fault tolerance in the manhole cover hidden danger detection industry. A data set of 13,420 manhole covers in various states is collected through network search, field investigation, etc., and after data cleaning, standardization, annotation, and partitioning, the YoSwin - Former model, which combines the advantages of YOLOv5 and Swin Transformer, is used for training. This model performs excellently in feature extraction and dealing with complex scenes, effectively improving the accuracy and efficiency of manhole cover hidden danger detection, and has far-reaching significance for ensuring the safety of urban residents' lives and property and promoting the intelligence of urban management.

Keywords: Manhole cover hidden danger detection, Intelligent identification technology, YoSwin - Former model, Data processing, Urban safety

1. Introduction

In the streets of the city, the manhole cover is like the "button" of the city. Although it is inconspicuous, it maintains the normal operation of the city. They undertake key functions such as drainage, power supply, communication, etc. They are an indispensable part of urban infrastructure and the guardians of urban image and security. However, in recent years, accidents involving people swallowing manhole covers have occurred frequently. These tragic events have not only brought indelible physical and mental trauma to the victims, but also shrouded the public's life and property safety like a haze, which has aroused great concern and profound reflection from all walks of life.

The traditional manhole cover management model has many disadvantages, such as multi-head management, separation of property rights and management rights, which leads to management confusion and repeated safety accidents. With the rapid development of urban construction, the role of manhole covers in modern smart cities is becoming more and more important. It is urgent to strengthen the safety management of manhole covers, which is not only related to the safety of public facilities, but also the key to protect the people's "safety under their feet" [1].

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/133/2025.20799

This project addresses high costs, detection challenges, and low fault tolerance in manhole cover hazard detection using intelligent identification technology. Real-time monitoring identifies potential risks like displacement, inclination, and breakage, triggering alarms for timely responses and preventing accidents. Intelligent manhole cover systems transmit data to control centers, enabling remote management and in-depth analysis of usage patterns and failure trends. This enhances maintenance efficiency, reduces safety incidents, and improves urban management. By ensuring public safety and enhancing life quality, the initiative promotes smarter cities and contributes to social stability [2].

2. Literature review

In the field of urban infrastructure management, the detection of manhole cover hidden dangers has become a research hotspot, aiming to improve urban safety and management efficiency.

Zhou B et al. studied the method of road manhole cover detection and classification based on smartphones [3]. They used the camera of smartphones to collect images and utilized deep learning algorithms to detect and classify the status of manhole covers. This method takes full advantage of the popularity of smartphones, making data collection more convenient and providing a low-cost and high-efficiency solution for manhole cover detection, especially suitable for large-scale daily inspection work.

Pasquet J et al. proposed a technique for detecting manhole covers in high-resolution aerial images of urban areas by combining two methods [4]. By integrating different image processing techniques, the accuracy of manhole cover detection is improved, especially in complex urban environments, with stronger recognition ability for manhole covers. This provides a reliable means for monitoring the status of urban manhole covers from a macroscopic perspective and helps urban planning and infrastructure management departments to timely grasp the distribution and status of manhole covers.

Liu W et al. used deep convolutional neural networks for detecting small manhole covers in remote sensing images, solving challenges like small target size and unclear features. This method improves large-scale urban infrastructure monitoring [5]. Liu H et al. focused on manhole cover detection in natural scenes by enhancing model adaptability to various environmental conditions, addressing issues like lighting changes and occlusions, ensuring stable detection [6]. This project integrates multiple technologies to optimize manhole cover hazard detection, improving accuracy, efficiency, and urban safety management, ultimately enhancing public safety.

3. Data collection and preprocessing

3.1. Data collection

The project data set is searched and abstracted by the network, and the original data set is formed by field investigation and photos. On this basis, the original data set is enhanced by Opencv, such as rotating and flipping, to optimize the data set effect. The total amount of manhole cover data is 13,420. According to the collection and the actual manhole cover situation, the data sets are divided into five categories: intact, damaged, missing, uncovered and problematic, and the number of data sets in each category is shown in the Table 1:

Table 1: Data set label classification

ID	Category	Name	Quantity/piece
0	Well cover intact	good	2633
1	Damaged manhole cover	broke	3133
2	Missing manhole cover	lose	2799

DOI: 10.54254/2755-2721/133/2025.20799

Table 1: (continued).

3	The manhole cover is not covered	uncovered	3016
4	Manhole cover problem	circle	1839

The data set of manhole covers used in this work is rich in categories, covering many scenes such as occlusion, telephoto, absence and shadow, and involving manhole covers of various shapes.

Figure 1: Data set partition

3.2. Data cleaning and standardization

To address noise in images, such as uneven lighting, shadows, and stains, filtering algorithms like median and Gaussian filtering can enhance picture quality. For blurry images that obscure key details, image definition evaluation algorithms assess clarity based on contrast and edge features, removing or processing unclear images. Duplicate detection algorithms, such as those using hash values, identify and eliminate redundant images, reducing data complexity and improving analysis accuracy. Images are standardized by resizing to uniform dimensions using scaling algorithms while preserving quality. To minimize color space variation, images are converted to a unified color space, like grayscale, or normalized to ensure channel consistency. File formats, naming rules, and metadata, such as labels and location information, are standardized for a cohesive dataset. Data augmentation techniques, including rotation, zooming, flipping, noise, and blur addition, simulate real-world conditions, expand the dataset, and enhance model performance.

4. Introduction to YoSwin-Former Model

4.1. YoSwin-Former:YOLO Combining Transformer Advantages

In this paper, the advantages of YOLOv5 and Swin Transformer are combined, which combines the real-time high precision of YOLOv5 in target detection with the ability of Swin Transformer in coding and capturing the global situation, thus achieving more efficient target detection performance. In this paper, the model is named YoSwin-Former [7].

YOLOv5 is a popular single-stage target detection algorithm known for its efficiency and accuracy, ideal for real-time applications. Swin Transformer, with its window and sliding window attention mechanisms, captures global and local image information efficiently. By embedding Swin Transformer Blocks into YOLOv5's Backbone, the model's feature extraction capabilities are enhanced. Optimizations in the Neck ensure real-time performance while maintaining high accuracy, combining YOLOv5's detection strengths with Swin Transformer's powerful feature extraction.

The benefits of this integration are remarkable. First of all, the global and local feature extraction ability of Swin Transformer can make up for the deficiency of YOLOv5 in feature representation [8],

DOI: 10.54254/2755-2721/133/2025.20799

thus improving the detection accuracy of the model. Secondly, because Swin Transformer adopts self-attention mechanism [9], it can better deal with complex relationships and contextual information in images, which is helpful to improve the model's ability to deal with complex scenes. Finally, by optimizing and adjusting the architecture and parameters of the model, the computational complexity and memory occupation can be reduced while maintaining the performance of the model, making the model more suitable for practical application.

4.2. YoSwin-Former Model network architecture

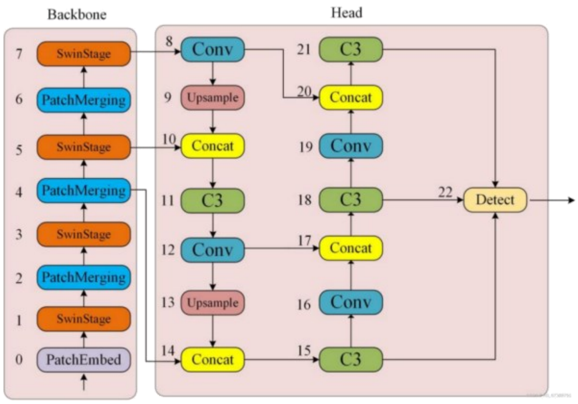

Figure 2: YoSwin-Former network architecture

The YoSwin-Former network integrates Swin Transformer’s feature extraction capabilities with YOLOv5’s efficient target detection. In the Backbone, the "SwinStage" module utilizes Swin Transformer’s Block, featuring convolution layers and window attention mechanisms for comprehensive feature extraction. These feature extractors are followed by convolution layers for further refinement. Additionally, up-sampling layers between convolution operations increase the feature map resolution, allowing the model to capture more detailed information. This combination of up-sampling and convolution enables the model to progressively extract features at multiple scales, enhancing its overall performance in target detection.

At the end of SwinStage, the outputs of different feature extractors are combined into a fully connected layer by Concat operation, and then output by a Sigmoid activation function [10]. Sigmoid function can limit the output value between 0 and 1, which is helpful for more accurate classification and detection of the model.

In the Head part, the network embeds input data into a feature space using a "Patch Embedded" layer, followed by multiple feature extractors (C3), convolution layers, and up-sampling layers to enhance feature extraction. Unlike the Backbone, the Head merges outputs from all feature extractors using Concat before applying the Sigmoid activation function, enabling better integration of multi-scale features. Finally, a detection layer determines the position and category of targets by utilizing the extracted features. By combining YOLOv5's robust detection framework with Swin Transformer's feature extraction capabilities, the network achieves superior performance in target detection tasks.

4.3. YoSwin-Former Model training design

YoSwin-Former algorithm model uses cross entropy loss function. Based on the concept of information theory, this function provides an effective way to measure the difference between the predicted results of the model and the actual labels. In the multi-classification problem, we usually have multiple possible categories, and each sample can only belong to one of them. The cross entropy loss function can effectively deal with this situation and help the model learn how to classify samples correctly.

The Adam optimizer is widely used in deep learning for large datasets, combining RMSProp and Momentum to calculate adaptive update steps. It adjusts each parameter's learning rate using first and second moment gradient estimates, enabling efficient parameter updates based on historical gradient information.

5. Model result analysis

When YoSwin-Former model reads image data from the folder [11], it visualizes the number of images in each category in the folder and the specific position of the manhole cover in each image and makes a summary diagram, as shown in Figure 3.

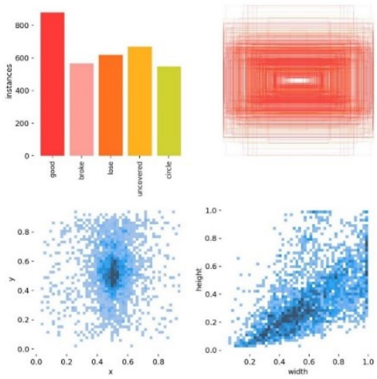

Figure 3: Image situation in folder

At the same time, the training results are visualized, and the trend of model accuracy transformation and model loss change is obtained. From the visualization results, it can be seen that with the increase of training times, the accuracy of the model is gradually improved and the loss is gradually reduced, showing an obvious optimization trend.

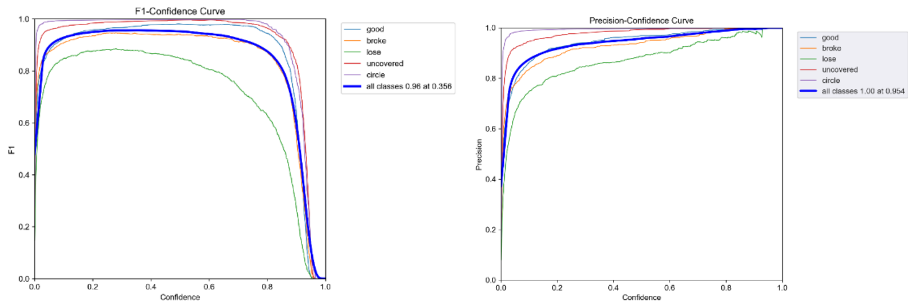

Figure 4: F1 and Precision index

F1 index in the process of model training is a comprehensive index to evaluate the performance of classification model. It comprehensively considers two key factors, Precision and Recall, aiming at finding a balance point to comprehensively evaluate the performance of the model.

In the process of model training, Precision is an important evaluation index of classification performance. It measures the proportion of samples predicted by the model as positive examples that are actually positive examples. In other words, the accuracy rate reflects the reliability or accuracy of the model prediction as a positive example. The higher the accuracy rate, the higher the proportion of positive samples predicted by the model, that is, the more accurate the prediction of the model.

DOI: 10.54254/2755-2721/133/2025.20799

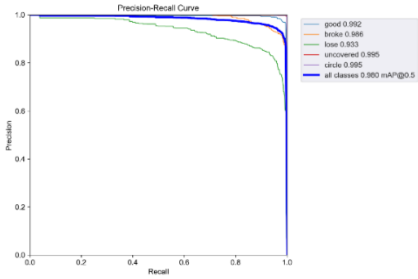

Figure 5: Recall index

Recall is an evaluation metric that measures the proportion of actual positive cases correctly identified by the model. Its range is [0, 1], with higher values indicating better performance in identifying positive cases. A low Recall suggests that the model misses many actual positive examples, leading to poor detection.

6. Conclusions

This study successfully constructed a manhole cover hidden danger detection system based on intelligent identification technology, effectively solving many problems in the traditional manhole cover management model. Through data collection and processing, a rich and standardized data set is obtained, laying a solid foundation for model training. The application of the YoSwin - Former model significantly improves the accuracy and efficiency of manhole cover hidden danger detection and shows good performance in practical application scenarios. This system can monitor the status of manhole covers in real-time, detect and warn of potential safety hazards in a timely manner, greatly improving the intelligent level of urban manhole cover management. Future research can further optimize the model and expand its application range to provide stronger technical support for the safety management of urban infrastructure.

References

[1] *Wang Yueyi, Zhang Han. Intelligent manhole cover data acquisition and transmission based on the Internet of Things [J]. Peasant Staff, 2020,(23):96+121.*

[2] *Ma Jiwei. Research on the design and management of smart terminals in "Sidun Technology" cities [D]. Shanghai University of Engineering Science, 2020.*

[3] *Zhou B, Zhao W, Guo W, et al. Smartphone-based road manhole cover detection and classification[J]. Automation in Construction, 2022, 140: 104344.*

[4] *Pasquet J, Desert T, Bartoli O, et al. Detection of manhole covers in high-resolution aerial images of urban areas by combining two methods[J]. IEEE Journal of selected topics in applied earth observations and remote sensing, 2016, 9(5): 1802-1807.*

[5] *Liu W, Cheng D, Yin P, et al. Small manhole cover detection in remote sensing imagery with deep convolutional neural networks[J]. ISPRS International Journal of Geo-Information, 2019, 8(1): 49.*

[6] *Liu H, Yan B, Wang W, et al. Manhole cover detection from natural scene based on imaging environment perception[J]. KSII Transactions on Internet and Information Systems (TIIS), 2019, 13(10): 5095-5111.*

[7] *Xu B ,Yu S .Improving Data Augmentation for YOLOv5 Using Enhanced Segment Anything Model[J].Applied Sciences,2024,14(5):*

[8] *Noor A ,Algrafi Z ,Alharbi B , et al.A Cloud-Based Ambulance Detection System Using YOLOv8 for Minimizing Ambulance Response Time[J].Applied Sciences,2024,14(6):*

[9] *Xu Y ,Wei H ,Lin M , et al.Transformers in computational visual media: A survey[J].Computational Visual Media,2022,8(01):33-62.*

[10] *Ma J ,Ouyang K ,Ma Z , et al.Transformer dense center network for liver tumor detection[J].Biomedical Signal Processing and Control,2024,91106066-.*

[11] *Li Jian, Du Jianqiang, Zhu Yanchen, et al. Overview of target detection algorithm based on Transformer [J]. Computer Engineering and Application, 2023,59(10):48-64.*