A victim risk identification model for nature-induced urban disaster emergency response

Fang, Weipeng; Reniers, Genserik; Zhou, Dan; Yin, Jian; Liu, Zhongmin

DOI
10.1111/risa.17456

Publication date
2024

Document Version
Final published version

Published in
Risk Analysis

Citation (APA)
Fang, W., Reniers, G., Zhou, D., Yin, J., & Liu, Z. (2024). A victim risk identification model for nature-induced urban disaster emergency response. *Risk Analysis*, *45*(3), 623-637. https://doi.org/10.1111/risa.17456

DOI: 10.1111/risa.17456

ORIGINAL ARTICLE

A victim risk identification model for nature-induced urban disaster emergency response

Weipeng Fang[1,2,3] | **Genserik Reniers**[3,4,5] | **Dan Zhou**[2] | **Jian Yin**[2] | **Zhongmin Liu**[1,2]

[1]School of Economics and Management, Tongji University, Shanghai, China

[2]The Institute of Disaster Medicine Engineering of Tongji University, Tongji University, Shanghai, China

[3]CEDON, KU Leuven, Campus Brussels, Brussels, Belgium

[4]Faculty of Technology, Policy and Management, Safety and Security Science Group (S3G), TU Delft, Delft, The Netherlands

[5]University of Antwerp, Antwerp, Belgium

Correspondence
Zhongmin Liu, The Institute of Disaster Medicine Engineering of Tongji University, 26th Floor, Zonghe Building (new), Shanghai East Hospital, Shanghai 200120, China.
Email: liu.zhongmin@tongji.edu.cn

Funding information
Tongji University, Grant/Award Number: 2023020046; Emergency Medicine and Critical Care, Grant/Award Number: PWYgf2021-03

Abstract

In recent years, nature-induced urban disasters in high-density modern cities in China have raised great concerns. The delayed and imprecise understanding of the real-time post-disaster situation made it difficult for the decision-makers to find a suitable emergency rescue plan. To this end, this study aims to facilitate the real-time performance and accuracy of on-site victim risk identification. In this article, we propose a victim identification model based on the You Only Look Once v7-W6 (YOLOv7-W6) algorithm. This model defines the "fall-down" pose as a key feature in identifying urgent victims from the perspective of disaster medicine rescue. The results demonstrate that this model performs superior accuracy (mAP@0.5, 0.960) and inference speed (5.1 ms) on the established disaster victim database compared to other state-of-the-art object detection algorithms. Finally, a case study is illustrated to show the practical utilization of this model in a real disaster rescue scenario. This study proposes an intelligent on-site victim risk identification approach, contributing significantly to government emergency decision-making and response.

KEYWORDS
disaster medicine rescue, nature-induced urban disaster, risk ranking, victim identification, YOLOv7-W6

1 | INTRODUCTION

With the rapid development of the economy and urbanization in China, most of the population is concentrated in high-density modern cities characterized by frequent human activities, commercial trade, and building constructions. As climate change progresses, the severity and frequency of nature-induced urban disasters are escalating, increasing the possibility of human exposure to disasters and resulting in higher casualty rates following each calamity (United Nations Office for Disaster Risk Reduction, 2023). Throughout the past decades, severe disasters, such as the Sichuan earthquake (2008), the Tianjin explosion (2015), the Global wildfires (2019), and the Henan floods (2021), have led to catastrophic casualties, presenting significant challenges for urban disaster medicine rescue. The report on the disaster medicine rescue works during the Armenian earthquake (1988) indicates that the survival rate of victims rescued within the first 3 h can reach 90%. However, if the rescue time extends beyond 6 h, the survival rate drops to 50% (Mao et al., 2017). The newly released national disaster prevention plan of China has also emphasized the expectation of efficient disaster medicine rescue (China National Commission for Disaster Reduction, 2022). These experiences highlight the importance of disaster medicine rescue in urban disaster emergency response to reduce casualties and minimize the disaster effects.

Disaster medicine rescue focuses on professional on-site rescue operations, including emergency medical treatment, disease prevention, and public health control (Emami et al., 2005; Fuse & Yokota, 2012). As a prerequisite, a precise understanding of the on-site situation and identifying the number of disaster victims is essential for decision-makers in formulating disaster medicine rescue plans. Nevertheless, rescue organizations always face challenges acquiring adequate information about the current situation in a short time window, leading to significant delays in rescue operations. Research also indicates that the primary cause of inadequate disaster responses is insufficient information rather than defective decision-making

(Lindell & Perry, 2012). In this context, on-site situation analysis, precise victim identification, and location are the most essential things of urban disaster medicine rescue.

Victim identification is a challenging task, particularly in unpredictable and unstructured environments. In the previous works, the rescuers merely relied on manual searches and remote sensing satellites to identify one or more physical parameters of victims trapped in disaster areas. These parameters can be voice, temperature, scent, motion, skin color, face, and body shape (De Cubber & Marton, 2009). However, some victims may be missed due to the limited rescue time and resources. Consequently, prompt remote data collection and intelligent victim identification have become promising tools for urban disaster medicine rescue tasks. Intelligent victim identification tasks rely on computer vision (CV) technologies, a subset of artificial intelligence (AI). It focuses on creating algorithms that enable computers to understand and interpret visual information from disaster zones (Iqbal et al., 2021). The CV technologies have been applied to disaster management, including disaster risk assessment, disaster detection, disaster evolution modeling, emergency evacuation simulation, and post-disaster damage analysis (Can et al., 2019; Li et al., 2023; Linh et al., 2021; Nag et al., 2020; Zhang et al., 2019). Nevertheless, the utilization of CV in disaster medicine rescue is quite rare, compared to its successful applications in other domains. In recent years, advancements in AI algorithms and hardware technologies have propelled the application of CV in victim identification tasks, particularly in using unmanned aerial vehicles (UAVs), global positioning system (GPS), and satellite imagery (SI) (Guha et al., 2022). Sun et al. (2016) proposed a comprehensive camera-based target detection and positioning system integrating various detection sensors for identifying wilderness victims. The identification and search algorithms equipped in the system enable real-time remote data collection and victim identification through an adaptive threshold target recognition algorithm. Andriluka et al. (2010) conducted a systematic comparative analysis of various victim detection algorithms, revealing that the most effective approaches involve flexible part-based representations and discriminatively trained detectors. Al-Kaff et al. (2019) introduced an algorithm for detecting and tracking humans in hash environments by analyzing aerial images. The concept combines skeleton pose and visual data, greatly enhancing detection accuracy. Jaradat and Valles (2020) provided a convolutional neural network (CNN)-based model that utilizes thermal infrared images with a positioning system to detect victims in harsh burning scenarios. Álvarez-Merino et al. (2022) integrated WiFi fine time measurement, ultra-wide band, and fusion technologies into an UAV to locate covered on-site victims. This method can search a wide range of disaster zones in a short time window, significantly improving the efficiency of search and rescue (SAR) tasks. Ulloa et al. (2023) also proposed a victim detection system based on deep CNNs using new multispectral images. The system has proven more efficient and precise in complex outdoor environments and weather conditions. Nevertheless, despite the successful application of CV in previous fields, the application of on-site victim identification for nature-induced urban disaster medicine rescue is a notable absence. Moreover, the detection models mentioned above typically employ simple background images for algorithm training, which may not be sufficient for complex urban disasters. In addition, early CNN-based algorithms always require massive model parameters, leading to increased power consumption and subsequently impacting detection results' real-time performance and accuracy. Therefore, in this article, we employed a more portable and efficient algorithm to identify disaster victims in real-time: the You Only Look Once (YOLO) algorithm. The YOLO series algorithms represent state-of-the-art target detection technology, striking a balance between high-quality inference speed and accuracy (Bochkovskiy et al., 2020; Li et al., 2022; Redmon et al., 2016; Redmon & Farhadi., 2017, 2018). The YOLOv7 algorithm has experienced improvements over its predecessors, resulting in significant enhancements in inference efficacy and accuracy (Wang et al., 2023). Nevertheless, unlike the single shot multibox sector (SSD) (Liu et al., 2016), the YOLOv7 algorithm can consider both speed and accuracy in detecting small targets, that is, its extended algorithm, the YOLOv7-W6 algorithm. Moreover, the one-stage algorithm may be slightly lower than the two-stage algorithm in terms of overall accuracy, such as the Faster R-CNN algorithm (Ren et al., 2015). Nevertheless, the two-stage algorithm will sacrifice more detection time and require more calculation load. It is worth noting that every algorithm has applicable scenarios, and choosing an algorithm based on several specific indicators is meaningless. In this context, for disaster victim detection, as we need to obtain the victim's real-time situation in disaster areas under a complex scene, the YOLOv7-W6 algorithm will have a better application prospect in terms of comprehensive performance.

This article first defines the key feature of the disaster victim when we need to understand the main pose of victims in the post-disaster situation. Subsequently, we collected relevant victim pictures from multiple resources and established a disaster victim dataset for model training and testing. After that, a comparative experiment is conducted to demonstrate the prior performance of the proposed model. Finally, a case study is illustrated to show the practical use of the YOLOv7-W6-based model. This study provides a robust method for victim identification in urban disaster medicine rescue, contributing significantly to government emergency decision-making. The rest of this article is organized as follows: Section 2 provides a detailed model description. The training and testing processes, and the results of the comparative experiment, are presented in Section 3. Section 4 illustrates a case study to show the model's practical use. Finally, the significance and potential applications of the algorithm are discussed in Sections 5 and 6, respectively.

2 | METHODOLOGIES

2.1 | Model design

According to the "Sort, Assess, Lifesaving Interventions, Treatment/Transport (SALT)" triage categories in disaster medicine, individuals who are seriously injured or unconscious often lose the ability to move freely and cannot escape from the disaster (Bhalla et al., 2015). These victims always need prompt disaster medicine rescue. The "fall-down" feature can be interpreted broadly to encompass any situation where a target experiences a significant loss of vertical stability. Therefore, the model established in this article concentrates on identifying victims classified as "fall down" during urban disasters. To clarify, it is worth noting that, in the following section, we will first apply a typical earthquake case to demonstrate the model's practicality in this article. However, the application scenarios of the recognition model include, but are not limited to, urban disasters, such as flooding, earthquakes, landslides, and even accidents where equipment or structures collapse, just like applying the YOLO series models to identify various objects in different pictures.

This model is divided into two modules: the input module and the recognition module. The input module receives files (pictures or videos) under the algorithm's requirements. In the case of video input, the video frames are sent to the recognition module. In real-world urban disaster scenarios, input files may be obtained from UAVs, SI, monitoring systems, and handheld devices. The recognition module is built on the YOLOv7-W6 algorithm. First of all, it conducts data enhancement and preprocessing operations on the input files. Second, it utilizes transfer learning and deep learning methods to extract features within the backbone network. A feature fusion operation is performed before the final output through detector heads. Finally, the model prints the identified victims and risk zones and further guides disaster medicine rescue operations. Figure 1 illustrates the framework of the on-site victim identification model.

2.2 | The YOLOv7-W6 algorithm

The YOLOv7-W6 algorithm employed in this study was specifically developed by the author for cloud GPU based on the original YOLOv7. This algorithm enhances the network's width and utilizes more channels, resulting in improved performance for small target detection in disaster zones (Wang et al., 2023). The algorithm is segmented into three parts: the input (Input), the feature extraction network (Backbone), and the detection head (Head).

In the Backbone, the SPPCSPC module fuses feature information from different layers. This module integrates the cross-stage partial network structure and the spatial pyramid pooling structure (He et al., 2015; Wang et al., 2020). Benefiting from the four distinct maximum pooling layers, this module can obtain different image features to extract large and small targets, enabling the algorithm to adapt to various resolution images. This structure also significantly reduces the computational load and enhances the algorithm's inference speed. Predictions are displayed on corresponding feature maps for targets of different sizes. Finally, the non-maximum suppression (NMS) algorithm is applied to eliminate highly redundant prediction boxes, obtaining the final output of four predictions. The detailed structure of the algorithm is presented in Figure 2.

The YOLOv7-W6 algorithm employs three loss functions: regression loss, confidence loss, and classification loss. The confidence loss and classification loss utilize the same BCEWithLogitsLoss function, whereas the regression loss employs the CIoU (complete intersection over union) loss function. The BCEWithLogitsLoss integrates a sigmoid activation function with the BCELoss (binary cross entropy loss) in a single layer, which is more stable compared to using a sigmoid function followed by a BCELoss function (Shi et al., 2020). Moreover, this approach also mitigates the vanishing and exploding gradient problems. The sigmoid activation function transforms the input into probability values within the range of 0–1 and can be described as

$$\sigma(x) = \frac{1}{1+e^{-x}}. \tag{1}$$

The BCEWithLogitsLoss is given by

$$L(x, y) = L = \text{mean}\{L_1, \ldots L_N\}^{\mathrm{T}}, \tag{2}$$

$$L_n = -\omega_n \left[y_n \cdot \log\sigma(x_n) + (1 - y_n) \cdot \log(1 - \sigma(x_n)) \right], \tag{3}$$

where N is the batch size, the reduction is not "none." n is the number of samples in the batch, and ω_n is a manual rescaling weight assigned to adjust the contribution of the loss for each element in the batch. y_n is the true label of the sample, and x_n is the raw output of the algorithm.

The CIoU loss is as follows (Zheng et al., 2020):

$$L\text{CIou} = 1 - \text{Iou} + \frac{p^2(b, b^{\text{gt}})}{c^2} + \alpha\nu, \tag{4}$$

$$\text{Iou} = \frac{|B \cap B^{\text{gt}}|}{|B \cup B^{\text{gt}}|}, \tag{5}$$

$$\alpha = \frac{\nu}{(1 - \text{Iou}) + \nu}, \tag{6}$$

$$\nu = \frac{4}{\pi^2}\left(\arctan\frac{\omega_{\text{gt}}}{h_{\text{gt}}} - \arctan\frac{\omega}{h}\right)^2. \tag{7}$$

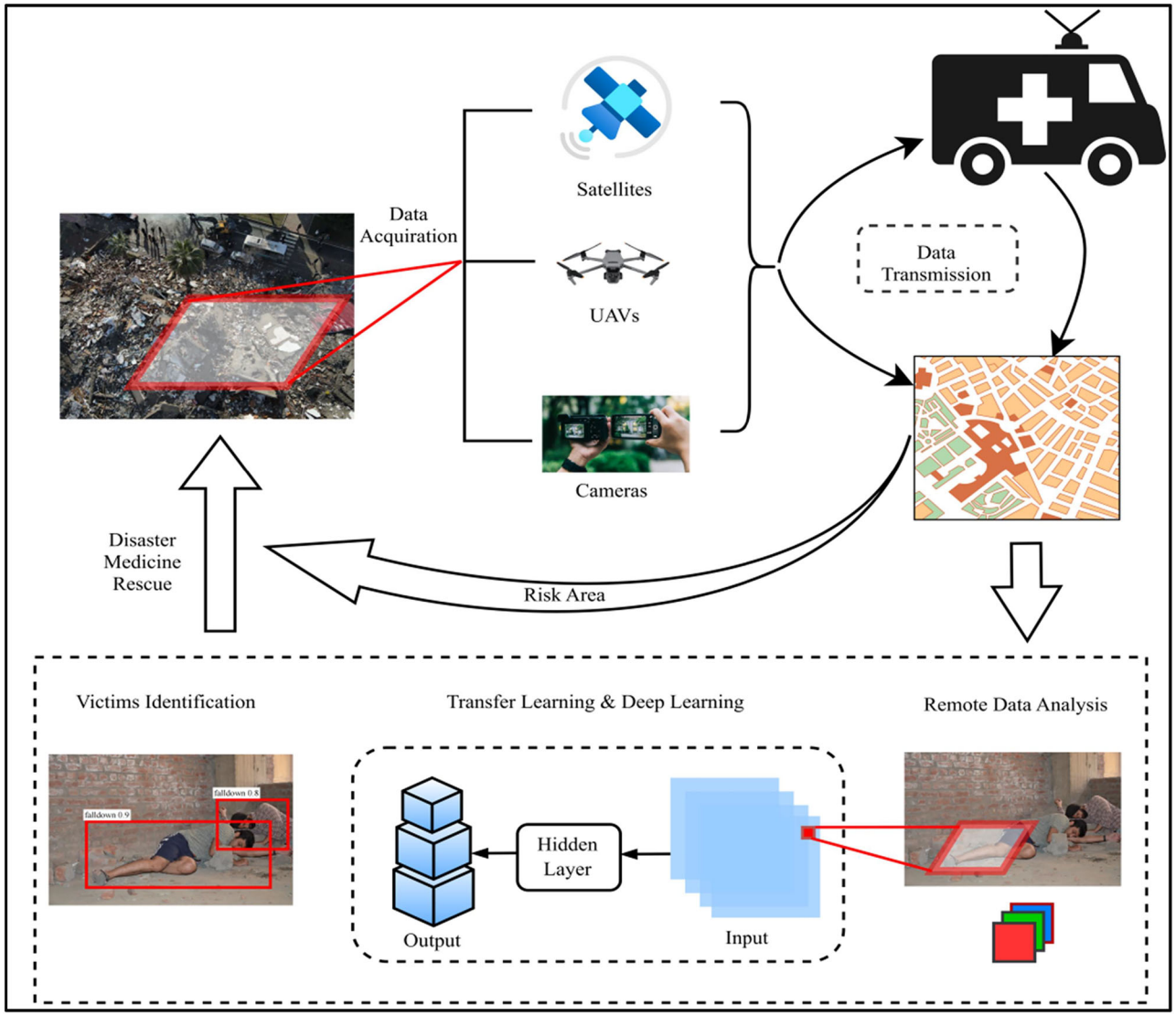

FIGURE 1 The framework of the on-site victim identification model based on the object detection algorithm. UAV, unmanned aerial vehicles.

The CIoU Loss is a loss function designed to optimize the calculation of intersection over union (IoU). The IoU represents the IoU between the predicted bounding box and the ground truth bounding box. p is the Euclidean distance between the centers of the true and predicted bounding boxes. b and b^{gt} are the center points of the true and predicted bounding boxes, respectively. c represents the diagonal length of the minimum enclosing rectangle that can simultaneously contain both the predicted and true bounding boxes. B^{gt} is the ground truth, and B represents the predicted box. α is a correction factor that considers the distance between the centers of the bounding boxes. ν is a term that considers the aspect ratio of the bounding boxes. ω_{gt} and h_{gt} are the width and height of the true bounding box, whereas ω and h are the width and height of the predicted bounding box.

3 | FINDINGS AND ANALYSIS: TRAINING AND TESTING

3.1 | Data collection and processing

As no specific open data are available for victim identification tasks during urban disasters, the original data used for training and testing were collected via multiple resources, such as official reports, disaster news, and the Internet. Most of the data were acquired by web crawlers from target websites, such as Baidu and Google (see Figure 3). The others are from official accident reports, for example, we can search for specific reports from the provincial and municipal emergency management bureaus on the Chinese government website. It is important to note that, as the model used in this article is

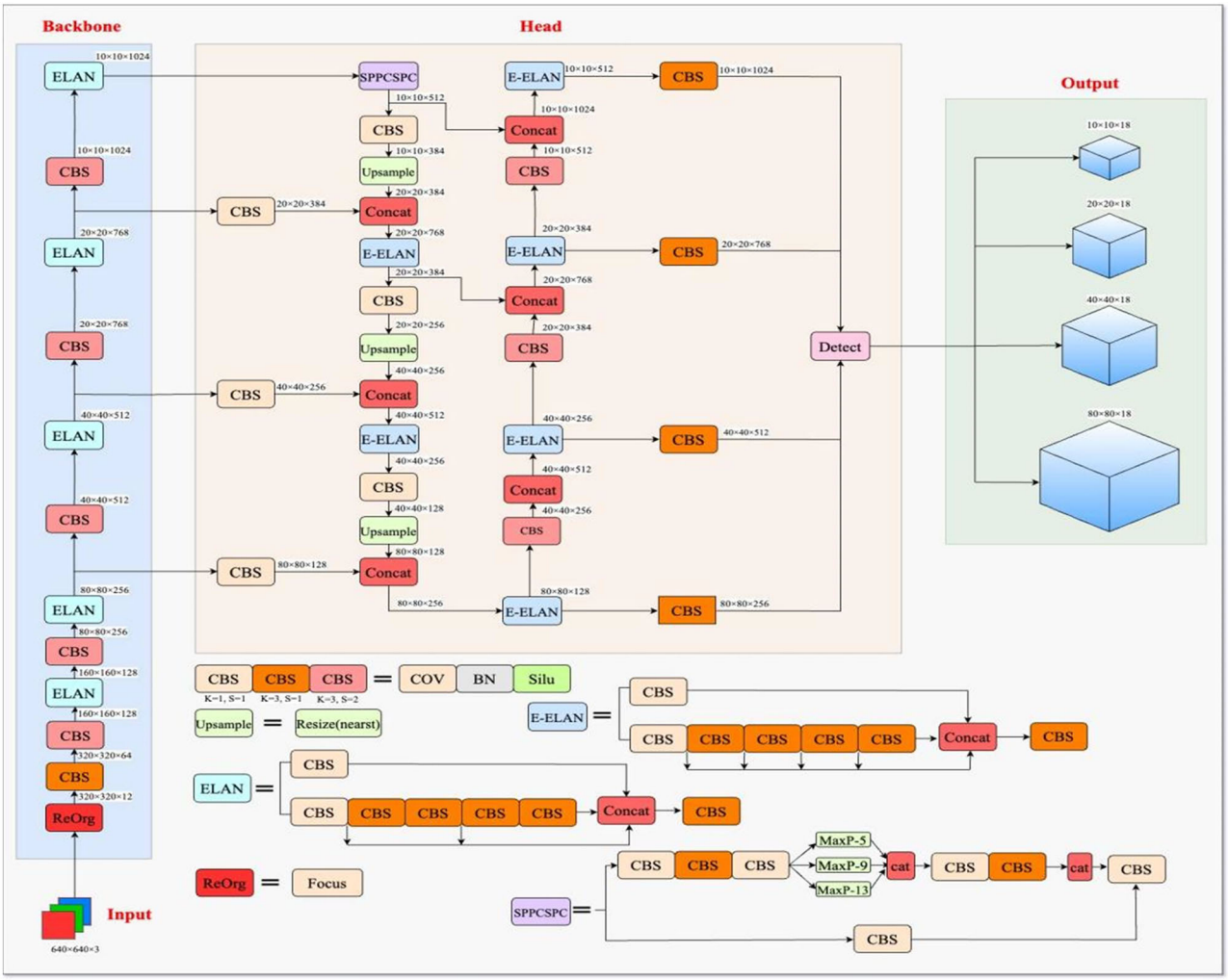

FIGURE 2 The detailed structure of the YOLOv7-W6 algorithm.

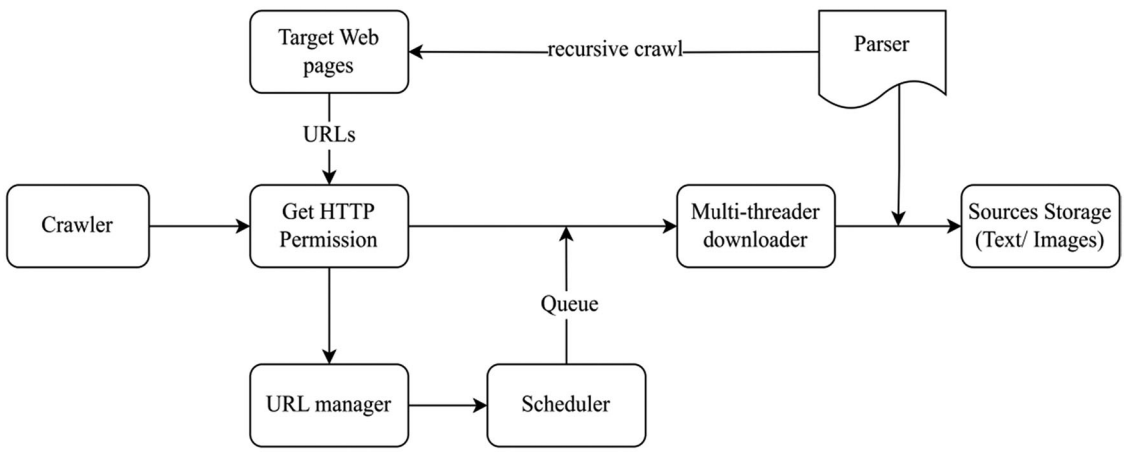

FIGURE 3 An example framework of web crawler process.

to identify "fall-down" victims, the background of the data is not strictly limited to urban disasters. The main reason is that the specific image of urban disaster victims is hard to obtain. Moreover, we found no recognition errors caused by scene changes in subsequent tests. On the contrary, we can improve the diversity of data. To this end, three principles should be followed when extracting target images: (i) The selected images should be taken during disasters or accidents, including fire, explosion, car crash, sports, stampede, flooding, and earthquake; (ii) the image should include at least one clear "fall-down" object; (iii) the image should show the location of the victim with minimal overlap.

Sequentially, we manually checked the rough data to ensure each image met the previous three criteria. After rigorous manual filtering and deduplication, we finally obtained a disaster victim dataset.

According to the results of the crawler, we will get a rough dataset, which may overlap, lack the target, or even be irrelevant to the topic of this article. Therefore, we compiled a clean and reliable dataset under a rigorous selection process. The detailed processes of filtering and deduplication are as follows:

1. Initial filtering: We first applied a basic principle to remove images that were clearly irrelevant. This included images with missing crucial information (e.g., victims, fall-down pose) or that were evidently false based on the recursive crawl.
2. Further exclusion: In this step, we further filtered out images based on the following criteria, including irrelevant entries, such as no victim appearing in urban disasters; incomplete data, images lacking essential information, such as there are no clear victims in the pictures, unsatisfactory pictures, the victims in the pictures have much overlap, or the pose of the victims is not "fall down."
3. Deduplication: This process would identify the same individual or scenario in multiple records. Each image is manually reviewed to remove duplicates. Furthermore, we were also determined whether the same victim was captured from different angles.

Sequentially, after filtering out irrelevant and incomplete images, we finally compiled a clean and reliable dataset of disaster victims. The dataset consists of 4576 pictures, further divided into a training set with 3203 pictures, a validation set with 915 images, and a test set with 458 images, following at 7:2:1. To meet the algorithm's requirements, we utilized the open-source LabelImg 1.4.0 tool for labeling the "fall-down" victims of all pictures (Everingham et al., 2015). An example of labeling a picture is illustrated in Figure 4. We can see that LabelImg uses a rectangular box, called a bonding box, to select the object with the class "fall down" displayed in the upper left corner. Sequentially, we can get the annotation information, which includes the class name, the center point of the bounding box, and the target object's width and height. We finally acquired the disaster victim dataset when we annotated all the pictures. It is worth noting that any researcher needs to strictly follow the regulations of dataset. The annotated dataset can be found at https://drive.google.com/file/d/1B-CrjB_EYJ0bJAguXwdNvAaOIen5Keuz/view?usp=drive_link.

To ensure accuracy and validate our labeling process, all annotators should be familiar with the software and the specific criteria for labeling "fall-down" victims. Moreover, every annotator must follow the guidelines to standardize the labeling process, including labeling principles and handling complex images. In this context, to ensure accuracy and rationality, we first invited several annotation teams to label the dataset and pass our consistency validation independently. Sequentially, we randomly selected a subset of labeled images by experienced annotators for a thorough review. Any discrepancies were discussed and resolved. Finally, we cross-validated the labeling results among different teams. It helped identify some systematic errors and improved the overall accuracy. After a comprehensive examination and adjustment, we have finally obtained the annotated disaster victim dataset.

3.2 | Equipment and parameter setting

The equipment and environment used for algorithm training and testing in this study are as follows: The CPU is a "12 vCPU Intel(R) Xeon(R) Platinum 8375C CPU @ 2.90GHz," the GPU is "RTX 3090 (24GB) * 1," and the deep learning framework PyTorch is employed with Python 3.8 version.

When training the algorithm, various data augmentation modules are applied for data preprocessing, including mosaic (Hao & Zhili, 2020), rotation (Islam et al., 2007), clipping (Ciocca et al., 2007), zooming (Shezaf et al., 2000), and MixUp (Zhang et al., 2017), as illustrated in Figure 5. Model parameters are configured as follows: The input image size is set to 640×640, the training epoch is 300, the initial learning rate is 0.01, the momentum parameter is 0.9, the weight decay parameter is 0.0005, and the batch size is 16. Additionally, the adaptive moment estimation algorithm is utilized to optimize the training process, and the early stopping method is implemented to monitor the loss of the training data (Bai et al., 2021; Kingma & Ba, 2014). The algorithm training process will be interrupted if the loss exhibits slow changes over an extended period.

3.3 | Metrics analysis

The metrics of the algorithm performance include precision (P), recall (R), and average accuracy (mAP). Precision (P) represents the ratio of true positives (TP) to the sum of TP and false positives (FP), whereas recall (R) denotes the ratio of TP to the sum of TP and false negatives (FN). The average accuracy (mAP) is a commonly used metric in object detection tasks. Higher values for these indicators indicate better algorithm performance. The calculations for P, R, and mAP

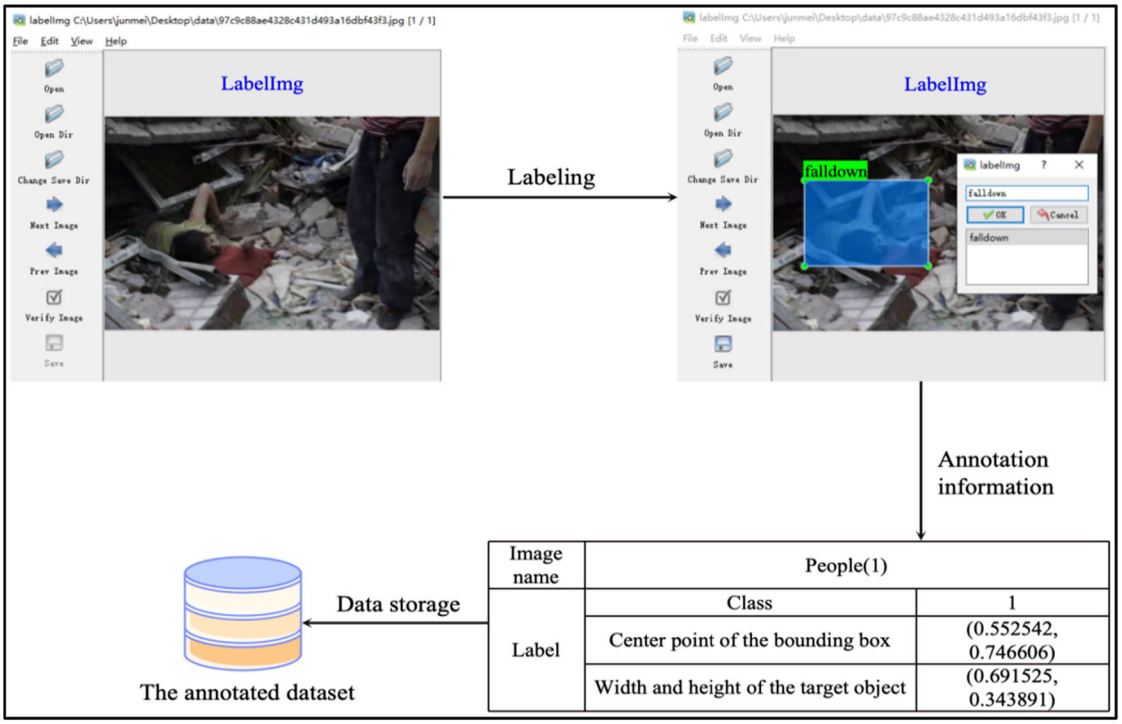

FIGURE 4 An example of labeling a picture using LabelImg.

FIGURE 5 Partial images of train dataset after preprocessing: (A) Mosaic: Randomly select four images from the training set and put them into a synthetic image, (B) Rotation: Randomly rotate the images, (C) Clipping: Randomly clip the image size, (D) Zooming: Randomly zoom in/out images to various scales and (E) MixUp: Randomly mix different images and lables to generate new training data.

are presented in the following equations:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{8}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{9}$$

where TP represent the samples correctly predicted as positive, and FP signify the samples incorrectly predicted as positive. FN indicate the samples incorrectly predicted as negative:

$$\mathrm{mAP} = \frac{\sum_{i=1}^{n} \mathrm{AP}_i}{n}, \tag{10}$$

where n represents the total number of classes, and AP_i represents the average precision metric explicitly calculated for the ith category among all the categories considered in our classification or detection task. Accordingly, in this article, the value of i is 1 because our goal is to identify a class of "fall-down" victims. It is important to note that this article employs the value of mAP@0.5. The mAP@0.5 measures the average precision when considering a predicted bounding box as correct if its IoU with the ground truth bounding box is greater than or equal to 0.5. The IoU is the intersection area divided by the union area of the two bounding boxes. A higher mAP@0.5 indicates better performance in detecting objects with sufficient overlap with the ground truth.

3.4 | Results

This study employs the transfer learning method to enhance the efficiency and performance of algorithm training and testing. This approach emphasizes transferring the features, representations, and algorithm parameters learned from one task to a new task, aiming to enhance the efficiency of the algorithm training process (Hosna et al., 2022).

In this study, the pre-training parameters are derived from the weights of a pre-trained algorithm on the COCO (common objects in context) dataset. The COCO dataset was developed by Microsoft and Cornell University; the dataset contains over 330,000 images (with 220,000 annotated images) and 80 categories, making it widely employed in various CV tasks (Lin et al., 2014). Applying the weights trained based on the COCO dataset as the initial parameters of YOLOv7-W6 can significantly improve the training efficiency and make it easier to obtain better weight parameters. Furthermore, a comparative experiment was conducted to assess the performance of different algorithms in identifying "fall-down" victims in urban disasters. The dataset was input to different algorithms on the same equipment and parameters. The results are presented in Table 1.

From Table 1, the YOLOv7-W6 and Faster R-CNN algorithms have better performance on on-site victim identification tasks than other algorithms during the training process, exhibiting improvements across all three indicators. For the Faster R-CNN, this two-stage algorithm performs best in the training process, aligning with previous expectations. For the YOLOv7-W6, the precision (P) reached 0.934, the best performance among the YOLO series of algorithms mentioned in this article. However, the model recall (R) is 0.907, indicating remarkably low-rate miss detection. Additionally, the mAP@0.5 increased significantly to 0.938, further demonstrating this algorithm's superiority. The detailed training results of the YOLOv7-W6 algorithm are illustrated in Figure 6, and the relationship between model precision and recall is depicted in Figure 7. The P–R curve is convex to the upper right, indicating the algorithm's stability and accuracy.

Table 2 presents the test results, also highlighting the superior performance of the YOLOv7-W6 algorithm in the victim identification task with the mAP@0.5 of 0.960, the precision (P) of 0.946, and the recall (R) of 0.919, respectively. Similar to the training results, the Faster R-CNN algorithm still performs better than others, with the highest mAP@0.5 and precision (P) value. In addition, all of these one-stage algorithms demonstrate efficient inference speed. The specific inference speed for a single image is less than 8 ms, whereas the NMS time is around 1 ms. The efficient structure of the CNN networks significantly contributes to this performance. In this aspect, the Faster R-CNN takes 25.6 ms to infer an image and 7.8 ms to find the optimal object bounding box, which is several times faster than the one-stage algorithms. Nevertheless, it is essential to note that YOLOv5x and SSD300 both employ a lighter structure with fewer parameters, resulting in the fastest image inference time, although the precision has declined. As mentioned above, the algorithm should be more balanced in practical rescue tasks. Therefore, the identification model based on the YOLOv7-W6 algorithm for disaster on-site victim identification tasks is an optimal selection.

Moreover, we carefully selected several test images in Figure 8 to visually demonstrate the mode's capabilities more intuitively. The results show that the model has high detection accuracy and fewer missed cases on the victim identification tasks. It further proves that the model is practical in the intelligent identification of on-site victims and contributes to disaster medicine rescue decision-making.

4 | CASE STUDY

This section uses a catastrophic earthquake disaster occurred in Sichuan, China, as an example of how this YOLOv7-w6-based model is applied to support decision-makers during disaster medicine rescue operations. The 2008 Sichuan earthquake (MW = 7.9) is one of the most devastating nature-induced urban disasters in the past decades, which caused more than 69,000 deaths and left 18,000 people missing. According to the report, the main causes of death were trauma and crush syndrome. Furthermore, there is also a significant death increase in respiratory infections, intestinal diseases,

TABLE 1 The training results of different object detection algorithms.

Algorithms	Precision (*P*)	Recall (*R*)	mAP@0.5
YOLOv5x	0.882	0.872	0.901
YOLOv7	0.810	0.819	0.850
YOLOv7-W6	0.934	0.907	0.938
YOLOv7-X	0.888	0.894	0.926
Faster R-CNN	0.956	0.93	0.941
SSD300	0.835	0.852	0.878

Abbreviation: CNN, convolutional neural network.

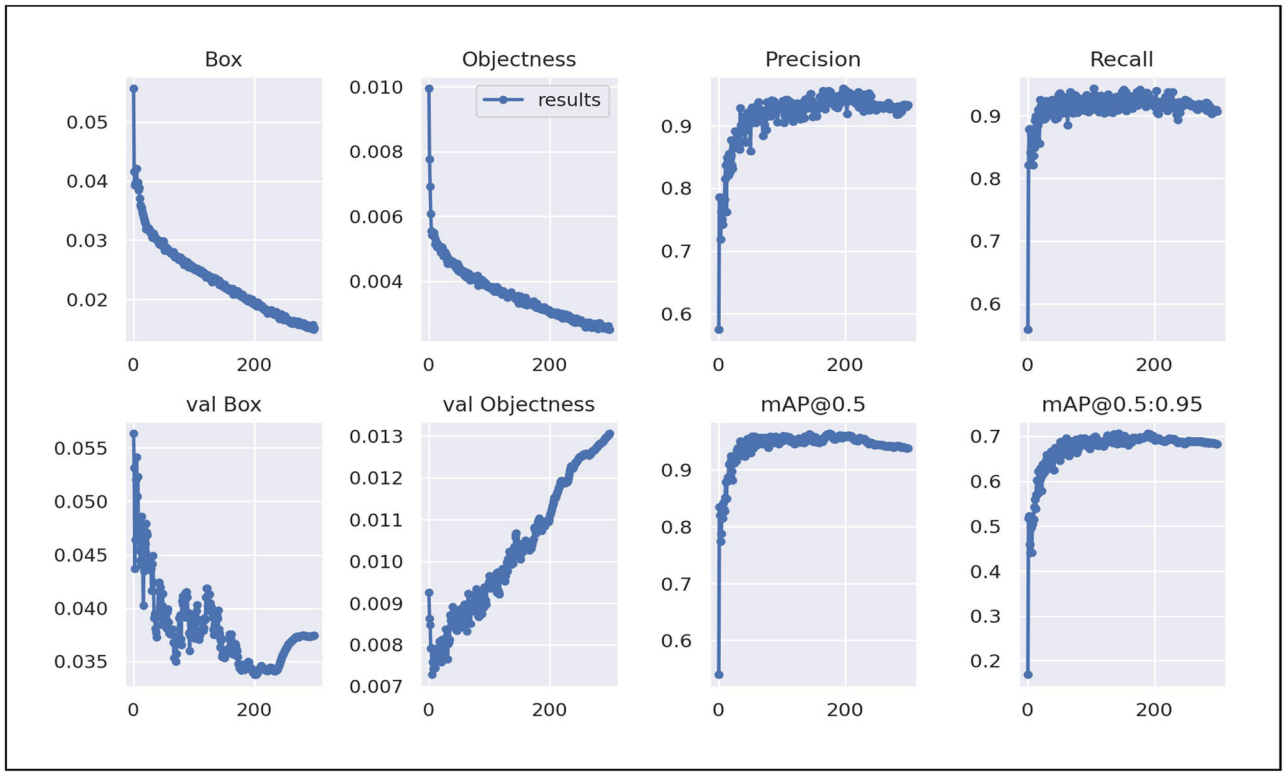

FIGURE 6 The training process of YOLOv7-w6.

and skin conditions in the following weeks after the disaster. Several months later, some survivors get into post-traumatic stress disorder problems (Zhou et al., 2019). This disaster has made us realize the importance of disaster medicine rescue in emergency response, and it is also a guideline to promote the development of disaster medicine in China.

4.1 | Data collection

According to the procedure we illustrated in the previous section, we first need to collect disaster pictures to identify the number of victims and their locations. Normally, we have several channels to acquire images and videos from disaster scenes, such as UAVs, on-site cameras, social media, and rescuers' feedback (Grimaz et al., 2022). Nevertheless, an earthquake disaster often causes other secondary disasters and ruins a city in a short time window; the power and communications systems may be briefly disrupted. In this context, the UAVs and on-site cameras are the most commonly used equipment to capture the exposed victims. Moreover, as the resources are limited, our rescue operations cannot cover all disaster areas; we only conducted the casualty analysis in a specific area, depending on the government's assignment (see Figure 9). The UAVs and on-site cameras are used to scan the area in hours, and these pictures are stored by the on-site database, for further identification. It should be noted that the on-site cameras might fail in some situations, leading to some victims being missed. Therefore, the UAVs need to scan the entire area. The UAVs' scanning area is circular, whereas the

TABLE 2 Performance of different algorithms on the RTX 3090.

Models	Precision (*P*)	Recall (*R*)	mAP@0.5	*t*0 (ms)	*t*1 (ms)
YOLOv5x	0.898	0.869	0.908	2.9	4.4
YOLOv7	0.804	0.813	0.853	6.0	6.8
YOLOv7-W6	0.946	0.919	0.960	5.1	6.0
YOLOv7-X	0.908	0.878	0.925	7.6	8.5
Faster R-CNN	0.956	0.903	0.964	21.6	29.4
SSD300	0.874	0.810	0.829	4.0	5.1

Abbreviation: CNN, convolutional neural network.

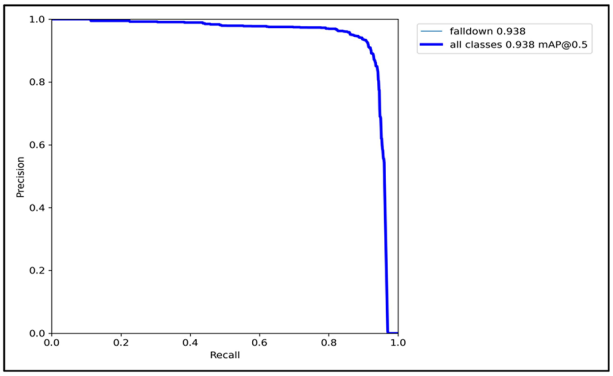

FIGURE 7 The relationship between precision (*P*) and recall (*R*).

on-site camera covers a fan-shaped area, the size of which is determined by the equipment. Once the entire area has been checked, we use the YOLOv7-w6-based model mentioned in the article to identify the victims in these areas.

In the following section, we selected some typical photos collected from the disaster scene as examples to see the identification results of this model. These pictures were taken from four points when the cameras and UAVs scanned the disaster area.

4.2 | Comparative analysis

In the previous section, we illustrated that the YOLOv7-w6-based model is optimal in accuracy and inference speed at the disaster victim dataset. Therefore, in this section, we applied the YOLOv7- and YOLOv7-w6-based models to test the performance of this model in a real-situation inference. The results are presented in Figures 10 and 11, respectively.

In Figure 10, we can find that both models can identify the victim under the collapsed concrete structure. However, we can see that in the results, the predicted bounding box in the right image cannot fully enclose the victim, even though it has a higher confidence score. This situation always occurs when the body is partially obscured, posing a challenge to the victim identification model. To this end, we used the YOLOv7-W6-based model to improve such cases. The YOLOv7-W6 algorithm adopts the "Reorg" module in the Backbone structure (Redmon & Farhadi, 2017). This module can increase the receptive field while reducing the computational load of the model, thereby improving the detection and identification capabilities of victims.

However, in Figure 11, we can see the results of multi-victim identification inference. The left image shows a significantly higher number of recognized victims than the right image, indicating that the model based on the YOLOv7-W6 algorithm demonstrates more reliable results. Moreover, we still manually verify the number of victims after algorithm analysis, only relying on manual correction when facing a large number of disaster photos is impractical. We would certainly select a more accurate victim identification model.

4.3 | Real-world victim identification and location

In this part, the utilization of the YOLOv7-W6-based victim identification model in a real disaster scene is presented in Figure 12. As mentioned above, the model can recognize disaster victims who are captured by UAVs or on-site cameras. In this context, after the disaster medicine rescue team enters the disaster zones (the circled yellow area of Figure 12), they will use numerous devices to collect images from the disaster zones. Casualty analysis will be conducted after the entire area has been scanned. Sequentially, the decision-makers will use the identified results to develop a reasonable disaster medicine rescue plan.

In this section, we take a typical scenario as an example. As shown in Figure 12, four points identify the victims, and we know the number of victims is 7. It is worth noting that this model can roughly identify the number of victims in the area; the reason is that in a massive disaster, there will be many potential victims trapped in the ruins, which are invisible. Therefore, the decision-maker can manually adjust the number of victims and calculate the risk level of the search area based on real-time disaster information. After obtaining the preliminary data on casualties at the disaster site, the decision-maker can calculate the risk level of the disaster area and decide the response action according to the requirements. The guidelines for risk ranking and the corresponding rescue operations are listed in Table 3. This is not a standard regulation. The risk ranking guideline in this article refers to the regulation of China (General Office of the State Council,

FIGURE 8 An illustration of partial test results with "fall-down" label.

FIGURE 9 The assigned rescue area and the location of the medical care base.

TABLE 3 The instruction of disaster medicine rescue operations.

Risk ranking	Description	Medicine rescue response
I (red)	Death > 30 or serious injured > 100	All team and medical resources
II (orange)	10 < death < 30 or 50 < serious injured < 100	All team and medical resources
III (yellow)	3 < death < 10 or 10 < serious injured < 50	All team, specific disaster medical resources, and a few other emergency medicine materials
IV (blue)	Death < 3 or serious injured < 10	Specific disaster rescue members and medical resources

FIGURE 10 Single-victim identification results of (A) the YOLOv7-W6-based model and (B) the YOLOv7-based model.

FIGURE 11 Multi-victim identification results of (A) the YOLOv7-W6-based model and (B) the YOLOv7-based model.

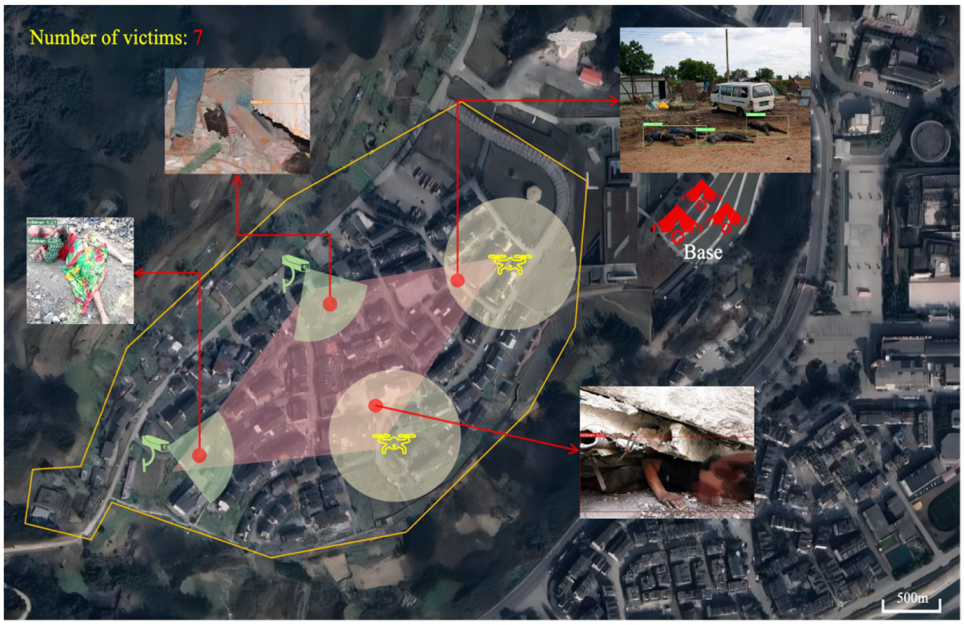

FIGURE 12 The results of real-time victim identification and location.

2016), and the corresponding operations are set according to disaster needs. This regulation can change according to the type of disaster. Accordingly, the decision-makers set the red area in Figure 12 as the key rescue area, especially the four points where the victims were found. At the same time, all team members and medical resources will be transported to these areas according to the disaster medicine rescue plan. Of course, if other victims are found, the decision-makers can dynamically adjust the rescue strategy to minimize the casualties of victims in the disaster area.

Moreover, we also summarize the different types of urban disasters and the main cause of death in these disasters, from the perspective of medicine, in Table 4. This statistical information provides a scientific reference for rescue workers to

TABLE 4 The main cause of death in nature-induced urban disasters.

Disaster category	The main cause of death
Earthquake	Trauma and crush syndrome
Flood	Apnea, trauma, hypothermia
Typhoon	Trauma
Fire	Apnea, poisoning, burn
Explosion	Burn and trauma
Toxic leakage	Poisoning and induced disease
Collapse	Trauma and crush syndrome
Stampede	Apnea and crush syndrome
Infectious disease	Inflammation, fluid and electrolyte imbalance, and induced disease
Disaster category	The main cause of death

carry more preferred supplies, depending on the type of disaster. In this context, the decision-makers can determine how to formulate a reasonable disaster medicine rescue plan based on these reasons and the number of on-site victims.

This case study only presents the initial phase of the disaster medicine rescue operation. Throughout the disaster rescue process, we will continue to make real-time adjustments to the plans based on the evolution of the disaster and the number of on-site victims. Additionally, the YOLOv7-W6-based victim identification model proposed in this article is a robust scientific basis for disaster medical rescue decision-making, which holds significant importance in mitigating the impact of nature-induced urban disasters.

5 | GENERAL DISCUSSION

In nature-induced urban disasters and accidents, the effective identification of on-site victims and formulating practical disaster medicine rescue plans are important for saving lives. With the development of CV technologies and deep learning algorithms, intelligent victim identification has rapidly developed, showing significant impacts in SAR, security monitoring, and other domains. However, due to the lack of open-labeled datasets and insufficient detection equipment, the improvement of the on-site victim identification model is still weak. Therefore, the victim identification task still requires a lot of human intervention, especially for the covered victims. However, if the on-site rescuers cannot provide clear input files, detection errors may occur frequently.

In the testing experiment of this study, the recall rate of the YOLOv7-W6-based model needs to be more satisfactory. This issue primarily arises due to the complex background and the presence of covered victims. Moreover, the probability of missing the target can be mitigated by providing images of the same victim from different angles. To this end, we can use UAVs or SI to take multiple photos of the same scene from different directions under such requirements. Furthermore, if possible, we can first use SI to scan the disaster area to grasp the overall situation of the disaster area, including road conditions, concentrated areas of personnel, and collapsed buildings. In addition, to increase rescue efficiency, we can use GPS to accurately locate the specific location of the victim, which significantly improves rescue efficiency and provides solid scientific support for further rescue route planning.

To make the practical use of the YOLOv7-W6-based victim identification model more apparent, we applied it to a real disaster rescue case. From the extended case study, we presented how this model combines images from the scene to identify victims, ranks the risk level, and formulates specific rescue actions based on the identification results. In addition, we will use various devices to scan the disaster area continuously during rescue operations. Whenever the number and location of victims are updated, the decision-makers will adjust the rescue strategy based on the latest information. Accordingly, every rescue task is also of great help in improving the accuracy of our model. In the future, when we have enough nature-induced urban disaster victim data, we will no longer rely on the existing database, making the model more generalizable.

In addition, due to the constraints of cloud service equipment, we observed that the response and data transmission speed were slow at the initiation period. This phenomenon leads to delays in real-time victim identification results. Therefore, if a dedicated server and a wireless local area network are established to replace the wide area network in urban disaster-affected zones, real-time detection performance would be significantly improved. Finally, although the on-site victim identification model helps formulate disaster medicine rescue strategies, estimating the number of victims remains meaningless without considering local population density and other social parameters. Therefore, the combination of local data and an intelligent victim identification model would be a promising method for precise victim estimation.

6 | CONCLUSIONS

In conclusion, the on-site victim identification model based on YOLOv7-W6 proposed in this study demonstrates

excellent detection results in various urban disaster scenarios. The model is expected to provide valuable technical support for disaster medicine rescue tasks.

In this study, we defined the critical feature "fall down" for urban disaster victim identification tasks based on the SALT triage principles and established an annotated dataset for model training and testing. The YOLOv7-W6-based model performance is superior in both inference speed and accuracy during model training and testing. In the training process, the model achieved a precision of 0.934, a recall of 0.907, and a mAP@0.5 of 0.938, indicating strong victim detection capabilities. In model testing, the corresponding values are 0.946, 0.919, and 0.960, demonstrating the generalization ability of this algorithm is good and it is also effective in the real-time identification of urban disaster victims. In addition, the comparison results highlighted the YOLOv7-W6 model's superior performance in detecting small targets compared to other algorithms. It is also important to note that the 5.1 ms inference time for a single picture and the 0.9 ms NMS time are essential for real-time detection in nature-induced urban disaster medicine rescue tasks. In summary, the YOLOv7-W6 algorithm is reasonable and balanced; we prefer a practical algorithm for real-world rescue tasks.

This study also has some limitations. The "fall-down" feature may not fully represent all urban disaster victims requiring immediate disaster medicine rescue. To better identify the victim, we may need more information and extract numerous features of the same target. Furthermore, the YOLOv7-W6-based model relies on clear victim pictures in the training and testing, providing a more critical obstacle to on-site data collection. Thus, in future work, we should focus on constructing a multi-source dataset to optimize model parameters and explore alternative methods to enhance the reliability of identification results.

ACKNOWLEDGMENTS

This work was funded by the International Exchange Program for Graduate Students, Tongji University, Grant/Award Number: 2023020046 and the Shanghai Pudong New Area Summit (Emergency Medicine and Critical Care) Construction Project, Grant/Award Number: PWYgf2021-03.

REFERENCES

Al-Kaff, A., Gómez-Silva, M. J., Moreno, F. M., De La Escalera, A., & Armingol, J. M. (2019). An appearance-based tracking algorithm for aerial search and rescue purposes. *Sensors*, *19*(3), 652.

Álvarez-Merino, C. S., Khatib, E. J., Luo-Chen, H. Q., & Barco, R. (2022). Victim detection and localization in emergencies. *Sensors*, *22*(21), 8433.

Andriluka, M., Schnitzspan, P., Meyer, J., Kohlbrecher, S., Petersen, K., Von Stryk, O., Roth, S., & Schiele, B. (2010). Vision based victim detection from unmanned aerial vehicles. In *2010 IEEE/RSJ international conference on intelligent robots and systems* (pp. 1740–1747). Taipei, Taiwan: IEEE.

Bai, Y., Yang, E., Han, B., Yang, Y., Li, J., Mao, Y., Niu, G., & Liu, T. (2021). Understanding and improving early stopping for learning with noisy labels. *Advances in Neural Information Processing Systems*, *34*, 24392–24403.

Bhalla, M. C., Frey, J., Rider, C., Nord, M., & Hegerhorst, M. (2015). Simple triage algorithm and rapid treatment and sort, assess, lifesaving, interventions, treatment, and transportation mass casualty triage methods for sensitivity, specificity, and predictive values. *The American Journal of Emergency Medicine*, *33*(11), 1687–1691.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Can, R., Kocaman, S., & Gokceoglu, C. (2019). A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality. *ISPRS International Journal of Geo-Information*, *8*(7), 300.

China National Commission for Disaster Reduction. (2022). *The 'fourteenth five-year' national disaster prevention and mitigation plan*. CNCDR. https://www.gov.cn/zhengce/zhengceku/2022-07/22/content_5702154.htm

Ciocca, G., Cusano, C., Gasparini, F., & Schettini, R. (2007). Self-adaptive image cropping for small displays. *IEEE Transactions on Consumer Electronics*, *53*(4), 1622–1627.

De Cubber, G., & Marton, G., (2009). Human victim detection. In *Third international workshop on robotics for risky interventions and environmental surveillance-maintenance* (pp. 1–9). RISE.

Emami, M. J., Tavakoli, A. R., Alemzadeh, H., Abdinejad, F., Shahcheraghi, G., Erfani, M. A., Mozafarian, K., Solooki, S., Rezazadeh, S., Ensafdaran, A., Nouraie, H., Jaberi, F. M., & Sharifian, M. (2005). Strategies in evaluation and management of Bam earthquake victims. *Prehospital and Disaster Medicine*, *20*(5), 327–330.

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, *111*, 98–136.

Fuse, A., & Yokota, H. (2012). Lessons learned from the Japan earthquake and tsunami, (2011). *Journal of Nippon Medical School*, *79*(4), 312–315.

General Office of the State Council. (2016). National natural disaster emergency rescue plan. *China Flood & Drought Management*, *1*, 29–34.

Grimaz, S., Malisan, P., & Pividori, A. (2022). Sharing the post-earthquake situation for emergency response management in transborder areas: The e-Atlas tool. *Journal of Safety Science and Resilience*, *3*, 72–86.

Guha, S., Jana, R. K., & Sanyal, M. K. (2022). Artificial neural network approaches for disaster management: A literature review (2010–2021). *International Journal of Disaster Risk Reduction*, *81*, 103276.

Hao, W., & Zhili, S. (2020). Improved mosaic: Algorithms for more complex images. *Journal of Physics: Conference Series*, *1684*(1), 012094. IOP Publishing.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1904–1916.

Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: A friendly introduction. *Journal of Big Data*, *9*(1), 102.

Iqbal, U., Perez, P., Li, W., & Barthelemy, J. (2021). How computer vision can facilitate flood management: A systematic review. *International Journal of Disaster Risk Reduction*, *53*, 102030.

Islam, S. M., Debnath, R., & Hossain, S. A., (2007). DWT based digital watermarking technique and its robustness on image rotation, scaling, JPEG compression, cropping and multiple watermarking. In *2007 international conference on information and communication technology* (pp. 246–249). IEEE.

Jaradat, F. B., & Valles, D., (2020). A victims detection approach for burning building sites using convolutional neural networks. In *2020 10th annual computing and communication workshop and conference (CCWC)* (pp. 0280–0286). IEEE.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y. F., Zhou, L. Y., Xu, X. M., Chu, X. X., Wei, X. M., & Wei, X. L. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976.

Li, J. F., Hu, Y. L., & Zou, W. G. (2023). Dynamic risk assessment of emergency evacuation in large public buildings: A case study. *International Journal of Disaster Risk Reduction*, *91*, 103659.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European conference, Zurich, Switzerland, September 6–12, 2014, proceedings, Part V 13* (pp. 740–755). Springer International Publishing.

Lindell, M. K., & Perry, R. W. (2012). The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis*, *32*(4), 616–632.

Linh, N. T. T., Ruigar, H., Golian, S., Bawoke, G. T., Gupta, V., Rahman, K. U., & Pham, Q. B. (2021). Flood prediction based on climatic signals using wavelet neural network. *Acta Geophysica*, *69*(4), 1413–1426.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *Computer vision–ECCV 2016: 14th European conference, Amsterdam, The Netherlands, October 11–14, 2016, proceedings, Part I 14* (pp. 21–37). Springer International Publishing.

Mao, M. X., Wang, L. X., Li, Q. L., Liang, H. P., & Cao, J. (2017). Expert consensus on disaster emergency and emergency response. *Chinese Journal of Hygiene Rescue (Electronic Edition)*, *3*(1), 1–11.

Nag, S., Pal, T., Basu, S., & Das Bit, S. (2020). CNN based approach for post disaster damage assessment. In *Proceedings of the 21st International Conference on Distributed Computing and Networking* (pp. 1–6). NY, USA: ACM.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788). IEEE.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263–7271). IEEE.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Ren, S. Q., He, K. M., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *1*, 91–99.

Shezaf, N., Abramov-Segal, H., Sutskover, I., & Bar-Sella, R. (2000). Adaptive low complexity algorithm for image zooming at fractional scaling ratio. In *21St IEEE Convention of the Electrical and Electronic Engineers in Israel. Proceedings (Cat. No. 00EX377)* (pp. 253–256). IEEE.

Shi, J., Li, Z., Zheng, H., Xu, Y., Xiao, T., Tan, W., Guo, X., Li, S., Yang, B., Xu, Z., Lin, R., Shangguan, Z. K., Zhao, Y., Wang, J. W., Sharma, R. H., Iyer, S. Y., Deshmukh, A., Mahalik, R., Singh, S., … Xu, C. L. (2020). Actor-action video classification CSC 249/449 Spring 2020 challenge report. arXiv preprint arXiv:2008.00141.

Sun, J., Li, B., Jiang, Y., & Wen, C. Y. (2016). A camera-based target detection and positioning UAV system for search and rescue (SAR) purposes. *Sensors*, *16*(11), 1778.

Ulloa, C. C., Garrido, L., Del Cerro, J., & Barrientos, A. (2023). Autonomous victim detection system based on deep learning and multispectral imagery. *Machine Learning: Science and Technology*, *4*(1), 015018.

United Nations Office for Disaster Risk Reduction. (2023). *GAR special report: Measuring resilience for the sustainable development goals*. United Nations Office for Disaster Risk Reduction. http://www.undrr.org/gar2023sr

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464–7475). IEEE.

Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 390–391). IEEE.

Zhang, G., Wang, M., & Liu, K. (2019). Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. *International Journal of Disaster Risk Science*, *10*, 386–403.

Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

Zheng, Z. H., Wang, P., Liu, W., Li, J. Z., Ye, R., & Ren, D. W. (2020). Distance-IoU loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. *34*, pp. 12993–13000). AAAI.

Zhou, L., Zhang, P., Zhang, Z. G., Fan, L. D., Tang, S., Hu, K. P., Xiao, N., & Li, S. G. (2019). A bibliometric profile of disaster medicine research from 2008 to 2017: A scientometric analysis. *Disaster Medicine and Public Health Preparedness*, *13*(2), 165–172.

How to cite this article: Fang, W., Reniers, G., Zhou, D., Yin, J., & Liu, Z. (2024). A victim risk identification model for nature-induced urban disaster emergency response. *Risk Analysis*, 1–15. https://doi.org/10.1111/risa.17456