Journal of Safety Science and Resilience 5 (2024) 355–364

Contents lists available at ScienceDirect

Journal of Safety Science and Resilience

journal homepage: www.keaipublishing.com/en/journals/journal-of-safety-science-and-resilience/

Human behaviour detection dataset (HBDset) using computer vision for evacuation safety and emergency management

Yifei Ding [a], Xinghao Chen [a], Zilong Wang [a], Yuxin Zhang [a,b,c,*], Xinyan Huang [a,*]

[a] Research Centre for Fire Safety Engineering, Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[b] State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China
[c] Department of Geotechnical Engineering, Tongji University, Shanghai, China

ARTICLE INFO

Keywords:
Image dataset
Object detection
Human behaviour
Public safety
Evacuation process

ABSTRACT

During emergency evacuation, it is crucial to accurately detect and classify different groups of evacuees based on their behaviours using computer vision. Traditional object detection models trained on standard image databases often fail to recognise individuals in specific groups such as the elderly, disabled individuals and pregnant women, who require additional assistance during emergencies. To address this limitation, this study proposes a novel image dataset called the Human Behaviour Detection Dataset (HBDset), specifically collected and annotated for public safety and emergency response purposes. This dataset contains eight types of human behaviour categories, i.e. the normal adult, child, holding a crutch, holding a baby, using a wheelchair, pregnant woman, lugging luggage and using a mobile phone. The dataset comprises more than 1,500 images collected from various public scenarios, with more than 2,900 bounding box annotations. The images were carefully selected, cleaned and subsequently manually annotated using the LabelImg tool. To demonstrate the effectiveness of the dataset, classical object detection algorithms were trained and tested based on the HBDset, and the average detection accuracy exceeds 90 %, highlighting the robustness and universality of the dataset. The developed open HBDset has the potential to enhance public safety, provide early disaster warnings and prioritise the needs of vulnerable individuals during emergency evacuation.

1. Introduction

For the past few decades, the increase in natural and man-made calamities such as earthquakes [1], building fires [2], floods [3] and stampede accidents [4] has promoted the urgent demand for public emergency safety research. Once an emergency occurs, prompt and proper evacuation is the key priority for human life safety. However, past disasters have illustrated that humans may lack knowledge of how or where to evacuate, while inefficient evacuation strategies and behaviours cause serious injuries and casualties [5,6]. Furthermore, people with special characteristics and behaviours and existing physical or mental troubles, such as pregnant women, elderly people, children or people with disabilities (Fig. 1), on emergency sites delay the evacuation process and increase the difficulty level of evacuation and rescue [7]. On the one hand, these special groups account for higher casualty risk, for example, the elderly aged over 65 comprise 32 % of home fire deaths but represent only 13 % of the population [8]. On the other hand, they may obstruct the egress of other evacuees owing to their slower speeds and the need for larger evacuation space [9]. Therefore, the provision of additional help and instruction for groups who need special attention is vitally significant to decrease injuries during disasters [10].

In recent years, there has been an increasing interest in applying artificial intelligence (AI) to the research and development of smart emergency management systems. Especially, visual object detection based on deep learning and computer vision has been critical and widely used in evacuation research and emergency system exploitation. For example, Zhao et al. [11] leveraged machine learning to investigate factors affecting pre-evacuation decision-making of building occupants. Huang et al. [12] applied computer vision algorithms to estimate crowd density and simulate evacuation aiming to reduce the stampede risk in public places. Cheng et al. [13] proposed a graph-based network to process real-time surveillance videos to detect and tally the number of evacuees in the target area for evacuation navigation. The above-mentioned studies illustrate the application potential of AI on evacuation, but they are not focused on vulnerable populations or multiple human behaviours.

* Corresponding authors.
E-mail addresses: yuxinzhang@tongji.edu.cn (Y. Zhang), xy.huang@polyu.edu.hk (X. Huang).

https://doi.org/10.1016/j.jnlssr.2024.04.002
Received 5 January 2024; Received in revised form 13 April 2024; Accepted 26 April 2024
Available online 2 June 2024

Object detection algorithms enable real-time and accurate detection of persons in the input frames and further extraction of people's distribution, movement speed, number of total evacuees and other physical information. In our previous study [14], the object detection model (YOLOv4) was redeveloped to build an evacuation monitoring system to extract and provide integrated information on the evacuation process; this system could only detect general human bodies. Because the vulnerable or people with disability groups as uncertainty in emergency response, the object detection systems not only should cater to the average occupants but also to groups that require more special attention. Therefore, accurate detection of various categories of evacuees and their diverse behaviours is valuable and helpful in guiding emergency evacuation and rescue during disasters. While most object detection systems show satisfactory performance for normal person detection, they cannot classify specific human behaviours.

To better identify the items, supervised learning, one of the most crucial branches of deep learning, is adopted and various models are developed for visual object detection. Generally, a well-trained object detection model requires a big dataset comprising numerous training samples and labels. However, most classical human image datasets such as the COCO dataset [15] and VOC Pascal dataset [16] overlook the diverse human categories and fail to distinguish them and label them differently. Therefore, only the mature AI model architecture is insufficient for complex human–behaviour identification, and a rich human dataset with vulnerable and other disruptive human–behaviour categories is necessary for developing a more powerful visual detection model.

In this paper, an open labelled Human Behaviour Detection dataset (HBDset) containing abundant specific human behaviour images and corresponding labels is introduced to help the public community to work on and enhance the effectiveness and accuracy of object detection algorithms for vulnerable people and other disruptive behaviour categories during evacuation. The proposed dataset contains common human behaviours during public emergencies such as those of pregnant women, children, people walking while playing on their phones and people using wheelchairs and crutches. Advanced object detection algorithms are adopted as examples to validate and calibrate the feasibility of the dataset. By establishing a more adequate database for human behaviours, this research provides an intelligent emergency management system framework for public emergency safety and early disaster warning and lays the foundation for the development of human recognition models. Fig. 2 illustrates the overall methodology of this work: (a) collecting raw image data, (b) classifying, annotating and splitting data to generate a detection dataset, (c) conducting experiments for the dataset using an object detection model and (d) providing perspectives of an intelligent monitoring system and a digital twin. The HBDset is released at https://github.com/JDmoric/HBDset-A_Human_Behaviour_Detection_Dataset.

2. Related work

2.1. Object detection algorithms

The advancement of computer vision has greatly contributed to the development and improvement of models for visual object detection. Over the past few years, various object detection algorithms have emerged as accurate and lightweight systems widely utilised in pedestrian recognition. The conventional object detection method is based on handcrafted feature extraction [17], e.g., the scale-invariant feature transform [18], shape contexts [19] and histogram of gradients [20].

With the booming development of deep learning, convolutional neural network (CNN)-based object detection models have gained tremendous popularity, such as R-CNN (region-based convolutional neural network) [21], R-FCN (region-based fully convolutional network) [22] and SSD (single shot multiBox detector) [23]. One notable algorithm is YOLO (you only look once) [24], which efficiently combines detection components into a single CNN operating on the entire image. By integrating localisation and classification tasks into a unified CNN framework, YOLO leverages features from the entire image to make predictions for each bounding box. Moreover, it can predict bounding boxes for multiple classes simultaneously, making global inferences about the entire image and its objects.

Up to now, the YOLO model has upgraded and iterated various classical versions with faster and more accurate detection performance, containing YOLOv2 [25], YOLOv3 [26], YOLOv4 [27] and YOLOv7 [28], as well as YOLOv5 [29] and YOLOv8 [30] presented by Ultralytics. The YOLO algorithm has demonstrated great potential in fire engineering when applied to fire detection. For example, Wang et al. combined YOLO and binocular cameras to realise real-time fire detection and fire distance estimation [31]. Moreover, the model structures of YOLO have the potential to be applied in human-behaviour detection and evacuation scenarios. For example, Li et al. [32] employed YOLO to extract the evacuees' movement parameters from the videos of an earthquake evacuation case to build an evacuation velocity–classification model. For human abnormal–behaviour detection, Ji et al. used T-TINY-YOLO and improved the network model with a CNN tailoring scheme [33]. Nguyen et al. presented a novel form of real-time human detection using smart video surveillance at the edge [34]. In the present study, we select YOLOv5, YOLOv7 and YOLOv8 to demonstrate the performance and feasibility of the proposed HBDset.

2.2. Human database

To train a specific AI model for an intelligent human detection

Fig. 1. Examples of special groups during evacuation: (a) children (The image is licensed under a Creative Commons License. Link: https://www.flickr.com/photos/giro555/10803144525/in/photostream/), and (b) a person with disability (The image is licensed under a Creative Commons License. Link: https://www.flickr.com/photos/undpeuropeandcis/5912638068).

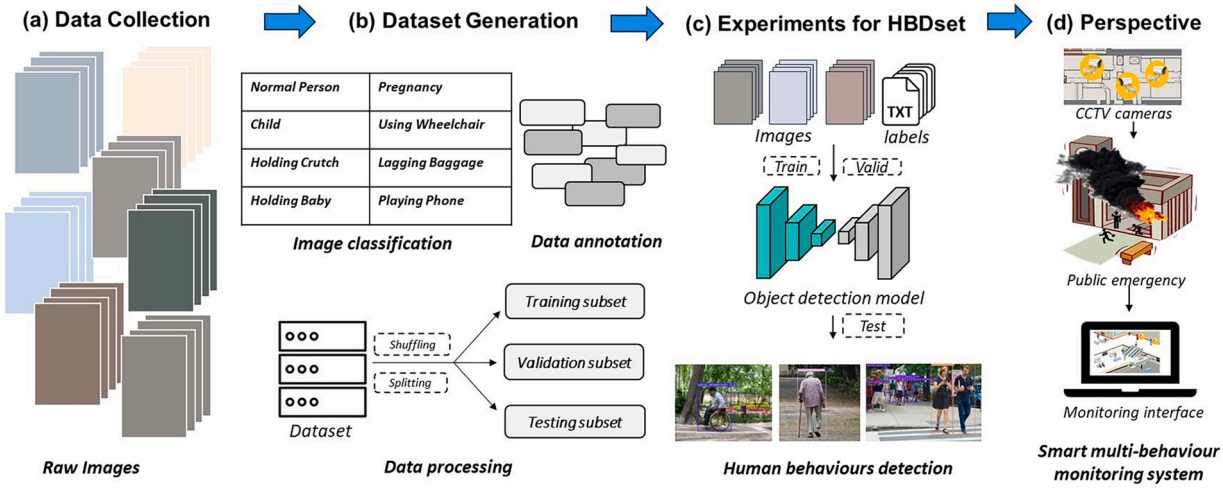

Fig. 2. Overall methodology of this work.

system, a well-annotated dataset with various human categories is the dominant component. Using human datasets, researchers and system developers customise the performance and function of the object detection model to well detect specific human classes. However, most of the popular datasets for object detection do not focus on the classification of different human behaviours. For example, even though PASCAL VOC (Visual Object Classes) datasets [16] annotate more than 10,000 instances of human and this number of MS COCO (Microsoft Common Objects in Context) datasets [15] is so many that more than 580,000 instances, the category of human has only 'person'. TDB [35] is a specified pedestrian dataset for supervised learning, but it has one 'individual pedestrian' category. Similarly, some human datasets contributing to person re-identification such as the CUHK dataset Market1501 [36] collect images of groups of people with different view angles and do not class other categories.

In addition, some studies contributed to human action recognition and established numerous corresponding human datasets with diverse human actions. For example, Gu et al. [37] proposed the Ava dataset that comprises 1.59 million action labels and densely annotates 80 atomic visual human actions such as 'sit', 'stand' and 'play instrument'. Some datasets classify human images using different attribute tags such as 'backpack', 'short hair', 'shorts', 'skirt', 'woman' and 'man', as well as mutual combinations. For example, the Richly Annotated Pedestrian (RAP) [38] dataset provides 84,928 images with 72 types of attributes and additional tags. Other such famous human attribute datasets include PA100-K [39] and PETA (PEdesTrian Attribute) [40].

Although these datasets classify human images into different attribute categories and make great contributions to person pattern recognition, the detection of overly diverse attribute tags, such as clothes, hairstyles and normal actions, is of little significance in terms of public safety in an emergency. Other researchers proposed road pedestrian datasets for autonomous driving and traffic safety. For instance, Zhang et al. [41] introduced a diverse city pedestrian dataset named CityPerson with 5000 images and four fine-grained person categories (pedestrian, rider, sitting and others). Notably, Sharma et al. [42] provided a new pedestrian dataset named BGVP (BG Vulnerable Pedestrian) focusing on vulnerable groups on a road, which contains 2000 images and 5932 bounding box annotations and classifies pedestrians into children without disability, elderly with and without disability and non-vulnerable people. Although the BGVP dataset tries to propose a guideline for vulnerable person identification, the other vulnerable groups such as pregnant women and groups with some special human behaviours such as playing on phones while walking or holding a baby must also be given increased attention in terms of emergency safety.

Although diverse and abundant human datasets have driven advances in computer vision–based person pattern recognition techniques, these datasets have neglected the importance of special human–behaviour detection in emergency scenarios. In this paper, a new open HBDset is introduced focusing more on vulnerable groups such as pregnant women, children, people using a wheelchair or a crutch and those with special behaviours that are not conducive during escape such as playing on phones while walking, holding a baby and lugging luggage. The comparison of the HBDset (groups with special behaviours are presented in bold) with existing human datasets is listed in Table 1. The introduction of the novel benchmark dataset will serve to advance the ongoing progress in the domain of specialised human group recognition, thereby stimulating heightened research engagement within this particular field.

3. Human behaviour detection dataset (HBDset)

3.1. Data classification

Vulnerable groups and people with disruptive behaviours for evacuation are considered to be special attention groups during evacuation scenarios. Once a public emergency occurs, the occupants should escape promptly and properly. However, individuals who belong to special attention groups, due to physical or mental conditions, may exhibit reduced responsiveness and mobility, along with an inadequate comprehension of evacuation strategies and routes. Therefore, these groups should be given extra attention, assistance or professional rescue support such as specialised wheelchairs or stretchers. An intelligent monitoring system should first encompass the initial identification, localisation and quantification of these demographic clusters. Naturally, rich detection datasets with abundant images of special groups are vitally significant for visual evacuee detection models based on deep learning. In accordance with the abovementioned evacuation monitoring demand, the evacuee detection instances in the HBDset are divided into eight categories: 'normal_person', 'child', 'holding_crutch', 'holding_baby', 'pregnancy' and 'using_wheelchair', 'lugging_luggage', and 'playing_phone'.

1) Normal_person: This group of occupants can rationally move and respond immediately once an emergency scenario occurs. These people do not have any visible physical disability or vulnerability and behave normally without carrying any items that may affect evacuation. In the HBDset, we assume that the instances of this group refer to adults who walk normally without any belongings or special movements.

Table 1
Comparing the HBDset with existing human datasets.

Type/Task	Datasets	Year	No. of human images	Human categories	Description
General	VOC [16]	2010	>100,000	1	For visual classification, detection and segmentation
	COCO [15]	2014	>580,000	1	/
Person re-id	CUHK01 [43]	2012	>1900	1	Focus on a person from various view angles
	Market1501 [36]	2015	>500,000	1	/
Action recognition	Ava [37]	2018	1.59 M	80	Human action: sit, kick, eat, stand and so on
Human attributes	PETA [40]	2014	19,000	61	Human attribute tags: gender, age, hairstyle, dress and so on
	RAP [38]	2016	84,928	72	/
	PA100-K [39]	2017		26	/
Road pedestrian	TDB [35]	2008	25,551	1	Pedestrian detection
	CityPerson [41]	2017	5000	4	Pedestrian, rider, sitting and others
	BGVP [42]	2022	2000	4	Vulnerable and disabled groups
Special groups	HBDset (this work)	2023	1523	8	Vulnerable and those with special human–behaviour groups

2) Child: This group refers to children without disability. Nonetheless, a child has limited understanding of the severity of an emergency scenario and weaker physical ability. Moreover, a child may not be familiar with the evacuation route and can easily become frightened and panicked during the escape, making them more vulnerable to injury. Therefore, children must be in the special attention groups and should be provided extra assistance and support during evacuation.
3) Holding_crutch: Most people in this group are elderly people or disabled. They have limited mobility and balance issues, making them more susceptible to falling or tripping during urgent evacuation. Moreover, they will find it very difficult to pass if they meet obstacles in the egress route. Therefore, focusing on their trajectory during evacuation and providing prompt assistance are critical.
4) Holding_baby: People who are holding a baby during an evacuation scenario are required to use at least one arm to hold the baby, and their visibility may be obstructed by the baby, making it harder for them to move and navigate quickly. At the same time, they need more energy to safeguard their and the baby's safety.
5) Pregnancy: Pregnant women have severe physical limitations, are more at risk of injury and are even prone to miscarriage. Hence, helping them evacuate or rescuing them is rather challenging, as it requires specialised strategies and labour assistance to guarantee their and the unborn baby's health as well as emotional support to help them stay calm.
6) Using_wheelchair: People using wheelchairs are highly likely to be elderly or disabled with limited mobility. The stress and panic during an emergency scenario can make it more difficult for them to remain calm and focused, which can further impede their ability to evacuate safely. Their dependence on wheelchairs makes it even harder for them to cross obstacles and stairs. Therefore, the monitoring system must detect and track them in real time and provide extra professional assistance and specialised rescue support.
7) Lugging luggage: This behaviour is disruptive for crowd evacuation. This group is not vulnerable or disabled, but their baggage is likely to make them move slowly and even inadvertently block the egress pathways or obstruct others' movement. Moreover, they may be more susceptible to falls, tripping or other accidents due to their luggage. Given these safety concerns, people who have lugging luggage during evacuation must also be given special attention.
8) Playing_phone: This behaviour is disruptive for crowd evacuation. Playing on phones during an evacuation scenario is very improper and dangerous. It may slow down the crowd evacuation process and put others at risk of injury. In addition, playing on phones can be a distraction and will prevent evacuees from being aware of their surroundings and the urgent atmosphere, making it harder for them to respond promptly to changing conditions or to follow instructions from emergency personnel. Hence, those who are playing on their phones must be detected, and on-time warnings must be provided.

In the process of constructing the database, we initially acquired data from well-established classical datasets. Notably, a portion of the 'normal_person' class images was sourced from the MS COCO dataset [10]. Furthermore, we drew inspiration from various classical image datasets, which draw their images from online resources. This inspiration led us to employ web crawlers and complementary plug-in components to systematically collect images from publicly accessible domains on the Internet. Additionally, a subset of the images incorporated into our dataset was captured within a public domain. In summary, the images of human behaviours are divided into eight categories based on vulnerable groups and improper behaviours.

3.2. Data annotation

In computer vision, the PASCAL VOC [16] format is widely used as the standard data annotation format in object detection–model training. The annotation works are conducted using the professional open-source image annotation software LabelImg [44] (Fig. 3). The labelled profiles are saved as XML files in the PASCAL VOC format, which contain key information regarding the images and annotations containing storage path, image resolution, class and object coordinates. The VOC format is the more general format for object detection–model training, while the recent versions of YOLO such as Versions 5–8 have used specialised TXT files in the YOLO format. It simplifies the annotation process using a single TXT file per image, which contains all the necessary information including the class label and bounding box coordinates of the objects. This streamlined format reduces the complexity of managing multiple XML files, and the two formats (PASCAL VOC format and VOC format) could be converted into each other using a Python programme. In this study, the annotation data are saved in the YOLO format. After data annotations, the results were inspected by other authors to ensure accuracy.

3.3. Data statistics

The HBDset comprises 1523 images with a total of 2923 objects, and each object was annotated using a bounding box and a label. On average, each image contains two bounding boxes. The image and object distributions are shown in Fig. 4, providing a quantitative understanding of the dataset. In the database, approximately 200 images were available for each category within the vulnerability groups, collectively constituting 13 %–14 % of the comprehensive database. By contrast, the total objects, which were substantially fewer in number, represented only 7.3 %–18 % of the entire database. Notably, the images depicting individuals classified as 'normal' comprised a mere 5.6 % of the total database, yet the targets associated with this category constituted a significant majority, encompassing 25 % of the comprehensive database. This observation aligns with real-world scenarios encountered within public spaces. The number of child objects is more than 500, achieving the highest ranking in all special attention groups. The category with the

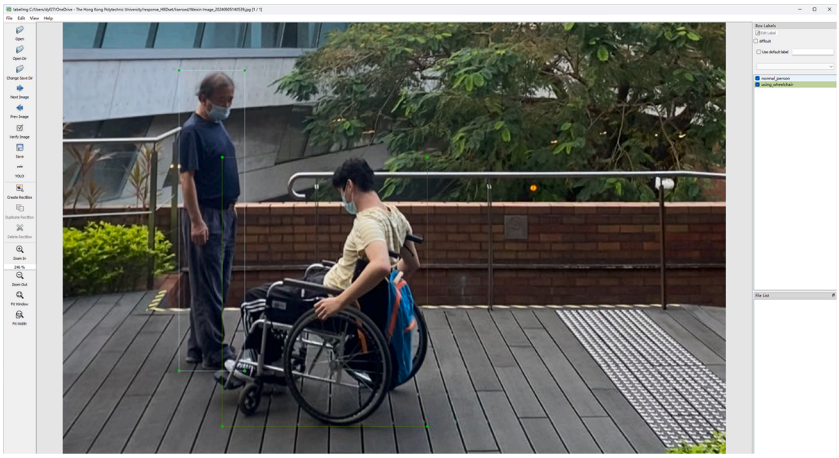

Fig. 3. Demonstration of data annotation by LablelImg (the original image is taken by authors).

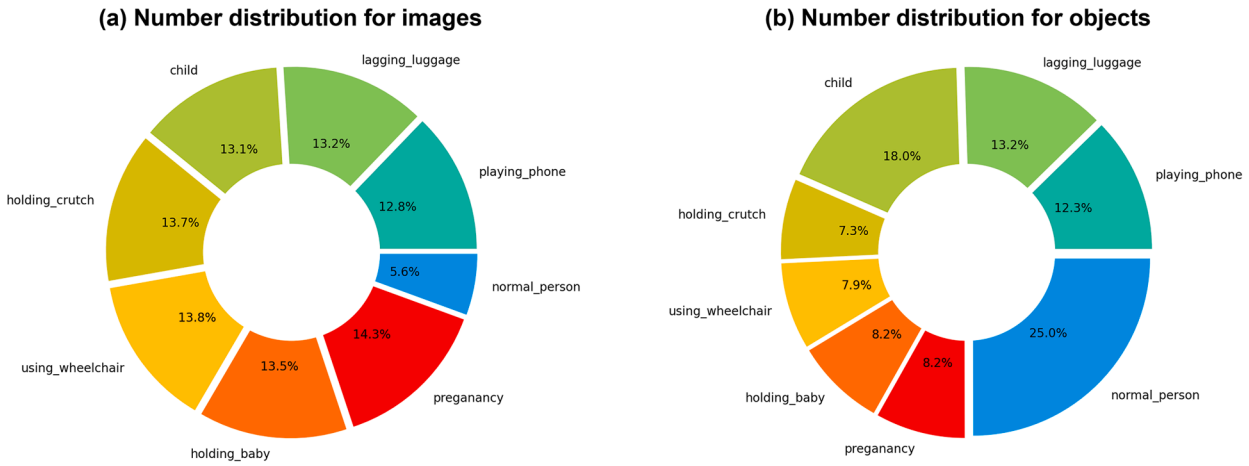

Fig. 4. Distribution of objects and images in each class: (a) images and (b) objects.

least number of objects is 'using_wheelchair' because there is only one object in most images of this group. More details of the dataset statistics are provided in Table 2.

4. Experiments for datasets

In this section, the object detection model is trained and tested using the proposed HBDset to present the feasibility of diverse human behaviour detection and provide a benchmark for relevant evaluation. In this study, the recent official versions of the YOLO families, namely, YOLOv5, YOLOv7 and YOLOv8, are selected to evaluate its performance on the HBDset. YOLOv7 optimises the model architecture and introduces many innovative techniques such as planned re-parameterised convolution to enhance the accuracy of real-time object detection.

Table 2
Image and object number of each category.

	Category	No. of images	No. of objects
No special attention	Normal_person	85	730
Vulnerable	Child	200	525
	Holding_crutch	208	214
	Holding_baby	206	239
	Using_wheelchair	210	231
	Pregnancy	218	239
Disruptive behaviour	Lugging_luggage	201	385
	Playing_phone	195	360

YOLOv7 achieves state-of-the-art performance and surpasses all known others on the MS COCO dataset. Moreover, YOLOv5 and YOLOv8 presented by Ultralytics are well packaged and achieve advanced robustness, which are widely used in academia and industry. Therefore, we select YOLOv5, YOLOv7 and YOLOv8 to conduct the experiment to validate the robustness and feasibility of our HBDset.

4.1. Model training

The colour transformation techniques including blur, darker, brighter and salt noise processing are used as the data augmentation method during model training. The colour transformation of images can increase image diversity to some extent, improve model generalisation for deep learning and reduce model over-fitting problems. Moreover, the augmented dataset has better robustness and reduces the interference of different backgrounds. Some instance images are shown in Fig. 5.

The images and annotations in the database were randomly disordered and then split into the training subset (1066 images, 70 %), validation subset (228 images, 15 %) and test subset (229 images, 15 %). The training set is utilised to train the model, the validation set is employed to assess the performance of the trained model during training and the test set is utilised to measure the accuracy of object detection.

The training process is conducted on a desktop computer with the following configuration: NVIDIA Geforce RTX 3070, 12th Gen Intel(R) Core (TM) i5–12490F 3.00 GHz. Pytorch [45] is applied as the deep learning framework to build the model structure. During training, the

Fig. 5. Demonstration of data augmentation (The raw unprocessed image is taken by authors).

batch size is set at 16 and a total of 300 epochs are set to make the model fully trained to convergence. To reduce training time and computational resources, the methodology of transfer learning [46] is applied to improve the performance of training results with the limited sample size. Transfer learning is a technique in deep learning in which parts of the knowledge learned from the original task are re-used to boost performance on a related similar task. The official YOLO version, as published, encompasses the original task of 'person' detection and the associated network weights. Consequently, the insights and knowledge acquired during the 'person' detection training phase can be effectively leveraged when training the model for the detection of various other human groups and behaviours. During the practical training process, transfer learning is executed by iteratively adjusting the weights while keeping the official weight profiles such as '*yolov7.pt*' frozen and retaining a portion of the neural network structure.

Fig. 6 illustrates the training loss and validation loss during the training process. The total loss adopted by YOLOv5 and YOLOv7 is accumulated in the loss of bounding box, loss of detection and loss of classification, shown as Eq. (1), while the total loss adopted by YOLOv8 is accumulated in the loss of bounding box, loss of classification and distribution focal loss, shown as Eq. (2). As the epoch increases, the training loss and validation loss steadily decrease until reaching a very low value. This indicates that the model has converged and achieved a consistent and optimal performance. The learning curve provides valuable insights into the robustness and generalisability of the dataset.

$$loss_1 = \lambda_1 \mathscr{L}_{bb} + \lambda_2 \mathscr{L}_{cls} + \lambda_3 \mathscr{L}_{obj}, \tag{1}$$

$$loss_2 = \lambda_1 \mathscr{L}_{bb} + \lambda_2 \mathscr{L}_{cls} + \lambda_3 \mathscr{L}_{dfl}, \tag{2}$$

where

$\mathscr{L}_{bb}$ represents the loss of the bounding box induced by pixel coordinates of the bounding box,

$\mathscr{L}_{cls}$ represents the loss of classification induced by the recognition of object category,

$\mathscr{L}_{obj}$ represents the loss of detection induced by the confidence for the inclusion of a target object,

$\mathscr{L}_{dfl}$ represents the distribution focal loss induced by the predicted bounding box offset and

λ_1, λ_2 and λ_3 are hyperparameters, and the setting values are listed in Table 3.

Another metric to evaluate the training performance of the deep learning object detection model is the mean average precision (mAP) [47]. mAP is a comprehensive evaluation index combining Recall (R) and Precision (P) shown in Eq. (3 and 4), which eliminates the limitation of using a single metric. Fig. 6 shows the average mAP curves of all classes in the training process, in which the blue solid curve represents the mAP score when the threshold equals 0.5, denoted as mAP @0.5, and the blue dash curve represents the average mAP scores in different thresholds ranging from 0.5 to 0.95, denoted as mAP @0.5:0.95. The mAP curves show that the prediction accuracy increases with the number of epochs and the average mAP score of all classes after 300 training epochs is >0.70. Overall, the three models afford stable training results, illustrating the high quality of the HBDset.

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{3}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{4}$$

Table 3
Hyperparameters of loss weights.

	YOLOv5	YOLOv7	YOLOv8
λ_1	0.05	0.05	7.5
λ_2	0.005	0.03	0.5
λ_3	1	0.7	1.5

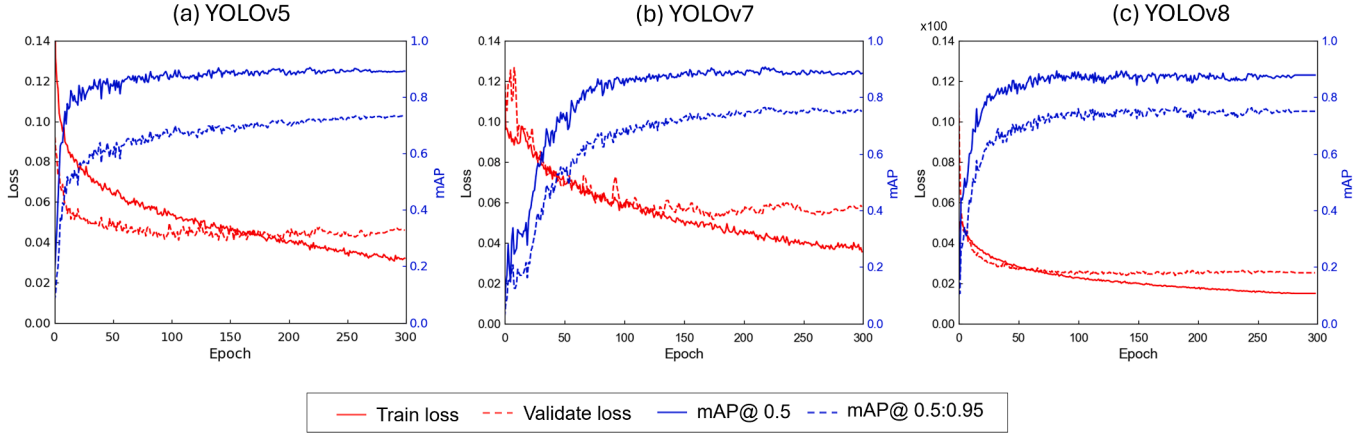

Fig. 6. Training process of the three object detection models on the HBDset: (a) YOLOv5, (b) YOLOv7 and (c) YOLOv8.

4.2. Model test

After training, the well-trained models were evaluated on the testing dataset containing a total of 229 images of eight categories. The Recall (R), Precision (P) and mAP@0.5 (mAP) were used as the metrics, and the testing results are shown in Table 4. The results show that the average mAP score of the three models is >0.89, where YOLOv7 achieves a higher mAP (0.91) score than YOLOv5 (0.86) and YOLOv8 (0.89). In addition, the category of 'using_wheelchair' achieves the highest mAP score, and the average mAP scores of the categories of 'holding_baby', 'holding_crutch' and 'pregnancy' are all >0.90. Notably, the class of 'normal_person' achieves the lowest accuracy score, which is significantly lower than the average mAP score. The attributes of each class are reasons for different classes achieving different mAP scores. For example, highly accurate classes such as including 'holding_baby' and 'using_wheelchair' have some very specialised characteristics such as the baby and the wheelchair, which make it much easier for the CNN to identify these attributes.

Moreover, the relevant images were randomly selected from the test dataset to demonstrate the detection performance of the well-trained models. The test dataset has never been used in the training process. We distribute the boxes of different human categories with various colours. A demonstration of some detection instances is shown in Fig. 7. In the demonstration, eight classes of human groups or human behaviours can be detected with high accuracy, and most of them achieve a confidence of >0.90. The results reveal that the detected human categories are accurately classified with few errors.

One of the limitations of this study is assuming that one individual has one specific behaviour or characteristic. However, some individuals may have several characteristics or show varying behaviours, which may lead to a more vulnerable situation. Therefore, the accurate detection and corresponding evacuation strategy of pedestrians with multi-attributes would be further researched.

4.3. Demonstration in Hong Kong international airport

To test the performance of the modified objection models trained by the HBDset, a demonstration in a public scenario of the Hong Kong International Airport is conducted for detecting and tracking special human groups, in which the tested videos are publicly available and collected from the Internet. In real evacuation scenarios, the types and distribution of people are often complex. To verify the robustness and universality of the HBDset and provide an evacuation scenario application case, we combined YOLOv7 and Deep Simple Online Real-time Tracking (DeepSORT) [48].

DeepSORT incorporates appearance information and employs Kalman filtering to track objects in an image. It employs a Hungarian algorithm to perform frame-by-frame association and quantifies through an association metric [49]. Therefore, this algorithm can be utilised to label personnel based on target detection, which will be beneficial to evacuation tracking and command. To simulate the surveillance perspective, strabismus was used for shooting experiments, and the target crossed multiple times during the tracking process. The result is shown in Fig. 8. The well-trained model by the HBDset can accurately complete the identification and the successful ratio of detection is >90 %, and the tracking performance is relatively stable, as shown in Videos S1 and S2 in the Supplementary Material.

5. Perspectives of intelligent monitoring and digital twin systems

The HBDset not only contributes to human recognition research but also helps develop the intelligent monitoring and digital twin systems. The HBDset could be used to train the object detection model and achieve automatic detection of vulnerable populations and other special behaviours, facilitating a more intelligent public safety system. Below is the discussion of the practical potential and contributions of the HBDset for the intelligent monitoring system and emergency management.

In public emergency disasters, intelligent monitoring should possess the capability to facilitate the precise localisation of occupants, monitor and record their movement trajectories and transfer the acquired monitoring data automatically. To realise these functions, an AI-based human detection algorithm is the critical prerequisite. Because evacuation behaviours in current public emergencies such as building fires or earthquakes are more diverse and complex, the system must detect multiple human groups or behaviours as well as consider various specific human safety scenarios during evacuation.

Furthermore, the intelligent digital twin system transferring the physical scenario as the virtual data integration would benefit public safety and emergency response. On the one hand, the digital twin system allows the monitoring of the trajectories and behaviour of various human groups, enabling the optimisation of evacuation strategies. On the other hand, it can integrate real-time processed information into the virtual interface, enhancing monitoring efficiency. To achieve this goal, the framework of an intelligent multi-behaviour digital twin system for public safety (Fig. 9) is proposed based on our HBDset. The flowchart of the proposed digital twin system contains three parts: (a) technical installation based on a CCTV (closed-circuit television) network, (b) AI engine based on a well-trained human-detection model and (c) monitoring user interface. These are responsible for raw video capturing, vision processing and bounding box generation, respectively, which can be employed to present detection results and issue corresponding instructions. For the implementation of the system, the CCTV network in public places can be used directly or more monitoring cameras can be installed as the eyes of the system.

A CCTV network captures raw videos in real time and synchronously transfers data to a cloud server or edge computing equipment. An AI engine would invoke the well-trained multi-behaviour detection model to process the received video signals and assign bounding boxes for detected humans. The detected objects would be marked with different coloured boxes based on the behaviour categories. In this step, the object detection model must be trained using the proposed HBDset before deploying the algorithm. In the future, we will add more top-view

Table 4
Testing results of the three models on the HBDset.

	YOLOv5			YOLOv7			YOLOv8		
Class	P	R	mAP	P	R	mAP	P	R	mAP
Normal_person	0.76	0.63	0.70	0.72	0.71	0.91	0.65	0.64	0.69
Child	0.87	0.79	0.90	0.80	0.83	0.84	0.90	0.70	0.86
Holding_crutch	0.84	0.90	0.93	0.86	0.99	0.99	0.97	0.91	0.97
Holding_baby	0.88	0.98	0.97	0.90	0.91	0.97	0.68	0.82	0.82
Lugging_luggage	0.77	0.69	0.77	0.84	0.85	0.87	0.91	0.92	0.93
Playing_phone	0.62	0.74	0.68	0.86	0.87	0.93	0.87	0.79	0.89
Pregnancy	0.99	0.79	0.96	0.83	0.97	0.91	0.92	0.80	0.94
Using_wheelchair	0.95	0.89	0.97	0.90	0.96	0.98	0.99	0.89	0.99
All	0.84	0.80	**0.86**	0.84	0.89	**0.91**	0.86	0.81	**0.89**

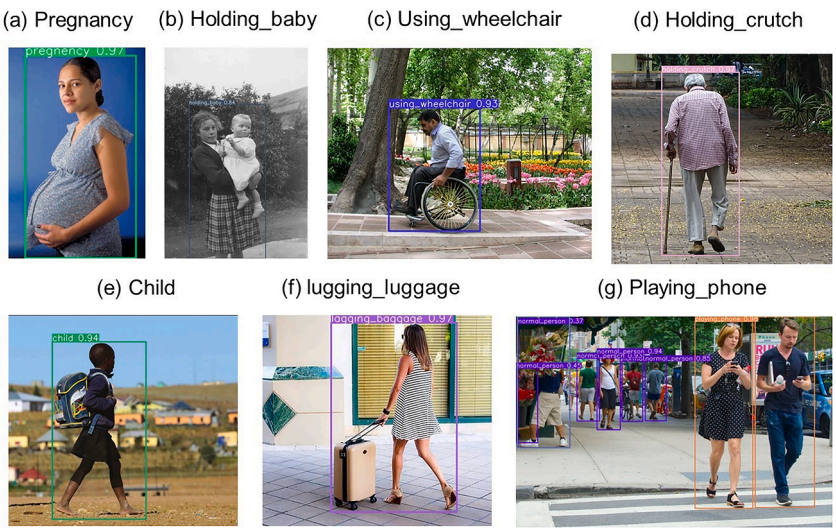

Fig. 7. Demonstration of detection effect in the test dataset. (The original undetected images are licensed under a Creative Commons License for free to share and adapt. Link: (a) https://courses.lumenlearning.com/suny-hccc-ss-152-1/chapter/pregnancy-and-childbirth/ (b) https://uhcl.recollect.co.nz/nodes/view/737 (c) https://www.tehrantimes.com/news/403689/Tehran-Municipality-to-enhance-services-to-people-with-disabilities (d) https://www.flickr.com/photos/joegoauk73/21314785534/in/photostream/ (e) https://www.news.uct.ac.za/article/-2017-06-16-the-state-of-sas-youth (f) https://gadgetsin.com/traxpack-stair-climbing-luggage-with-bluetooth-tracker-power-bank-and-more.htm (g) https://www.flickr.com/photos/yourdon/15077549298/in/photostream/).

Fig. 8. Detection and tracking effect in the Hong Kong International Airport (copyright: authors).

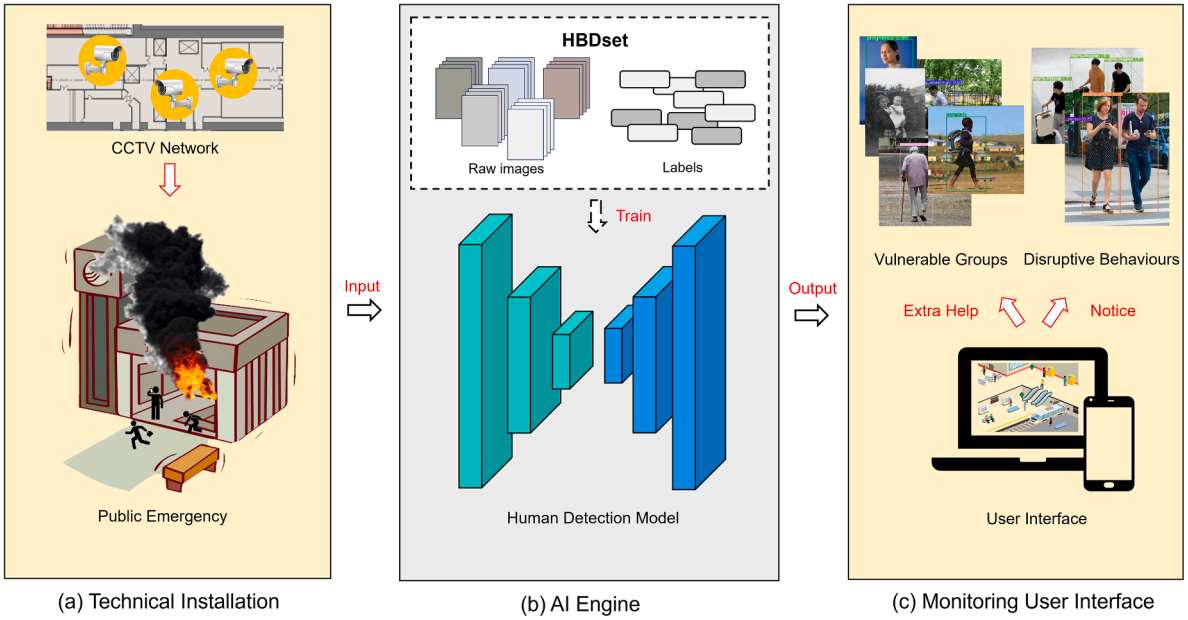

Fig. 9. Flowchart of an intelligent digital twin system for monitoring public safety and improving evacuation during earthquakes, building fires and stampedes.

images to this dataset to further improve the model performance. The third part is the output edge to perform the processed information in the user interface for on-site or remote evacuation directors. The displayed relevant information should contain the real-time locations, trajectories and counts of different human groups. The intelligent digital twin system enables providing commands and advice based on the information of detected vulnerable groups and people with improper behaviours. For example, the system would provide suggestions that more than five ambulance stretchers are needed or that one pregnant woman and two women holding babies need extra help. Meanwhile, during the escape, the system would warn people playing on their phones or people with lugging baggage via alarms or on-site instructors to avoid falling or even causing a stampede.

The core component of the proposed intelligent system is the modified object detection model pre-trained using the corresponding human image dataset. Therefore, the presented HBDset lays the foundation for developing the described intelligent system, which is the major contribution of this study. Building a practical intelligent digital twin system for use in public emergency would need the resolution of a multitude of engineering challenges, such as whether the latency of data transmission caused by the limits of wireless devices and computation resources affects the system performance of real-time instruction and feedback. Moreover, the artificial factor undertakes much major work in the intelligent system instead of a fully unmanned operation. Nonetheless, the automatic recognition and detection of diverse human behaviours is the first step to building a fully unmanned and high-level smart monitoring system framework in the future.

6. Conclusions

In this paper, an open human behaviour–detection dataset for deep learning in public emergency safety is proposed, denoted as the HBDset. The HBDset collected and annotated more than 1500 images with more than 2900 object bounding boxes, containing diverse vulnerable human groups or people with improper behaviours during evacuations. The collected images were augmented to increase the generalisation and annotated as standard object detection format for deep learning. The continued refinement of classical object detection models for complex human behaviour detection, resulting in the attainment of significant levels of accuracy in vulnerability group detection, serves to underscore the suitability of the dataset for the advancement of human behaviour detection models. Furthermore, a comprehensive framework for an intelligent multi-behaviour monitoring system is expounded, with potential applications in the domain of public emergency management and early disaster alerts.

In summary, by constructing this repository and the envisioned intelligent digital twin system for monitoring human conduct and public safety, we aspire to draw increased focus towards susceptible demographics during disaster evacuation. Through the identification of spatial distribution, quantification and trajectory of special human groups, the consideration afforded to the human behaviours can be augmented, thus enhancing the likelihood of secure egress from disaster-stricken areas. Additionally, such data will empower decision-makers to formulate more efficacious evacuation strategies, which will facilitate the entire evacuation process and public emergency response.

CRediT authorship contribution statement

Yifei Ding: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Xinghao Chen:** Writing – original draft, Investigation, Formal analysis. **Zilong Wang:** Writing – review & editing, Supervision, Conceptualization. **Yuxin Zhang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Xinyan Huang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is funded by the Hong Kong Research Grants Council Theme-based Research Scheme (T22-505/19-N), the National Natural Science Foundation of China (52204232) and MTR Research Fund (PTU-23005).

Data availability

The HBDset uses the images for non-commercial research and/or educational purposes.

HBDset with images and annotations, the Python script of splitting dataset into train, valid and test subset, well-trained weights of customed YOLOv5, YOLOv7 and YOLOv8 for human behaviour detection and corresponding code repository links are available on GitHub (https://github.com/JDmoric/HBDset-A_Human_-Behaviour_Detection_Dataset). The file README.md plays the role of guidance of HBDset for the users.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jnlssr.2024.04.002.

References

[1] S. Grimaz, P. Malisan, A. Pividori, Sharing the post-earthquake situation for emergency response management in transborder areas: the e-Atlas tool, J. Saf. Sci. Resil. 3 (1) (2022) 72–86, https://doi.org/10.1016/j.jnlssr.2021.12.001.
[2] Y. Zhang, X. Zhang, X. Huang, Design a safe firefighting time (SFT) for major fire disaster emergency response, Int. J. Disaster Risk Reduct. 88 (2023) 103606.
[3] A.F. Lee, A.V. Saenz, Y. Kawata, On the calibration of the parameters governing the PWRI distributed hydrological model for flood prediction, J. Saf. Sci. Resil. 1 (2) (2020) 80–90, https://doi.org/10.1016/j.jnlssr.2020.06.006.
[4] X. Hu, H. Zhao, Y. Bai, J. Wu, Risk analysis of stampede in sporting venues based on catastrophe theory and Bayesian network, Int. J. Disaster Risk Reduct. 78 (2022) 103111, https://doi.org/10.1016/j.ijdrr.2022.103111.
[5] R. Lovreglio, D. Borri, L. dell'Olio, A. Ibeas, A discrete choice model based on random utilities for exit choice in emergency evacuations, Saf. Sci. 62 (2014) 418–426, https://doi.org/10.1016/j.ssci.2013.10.004.
[6] R. Lovreglio, A. Fonzone, L. dell'Olio, D. Borri, A study of herding behaviour in exit choice during emergencies based on random utility theory, Saf. Sci. 82 (2016) 421–431, https://doi.org/10.1016/j.ssci.2015.10.015.
[7] W. Weng, J. Wang, L. Shen, Y. Song, Review of analyses on crowd-gathering risk and its evaluation methods, J. Saf. Sci. Resil. 4 (1) (2023) 93–107, https://doi.org/10.1016/j.jnlssr.2022.10.004.
[8] S.W. Gilbert, D.T. Butry, Identifying vulnerable populations to death and injuries from residential fires, Inj. Prev. 24 (5) (2018) 358–364.
[9] J. Koo, Y.S. Kim, B.-I. Kim, Estimating the impact of residents with disabilities on the evacuation in a high-rise building: a simulation study, Simul. Model. Pract. Theory 24 (2012) 71–83.
[10] J. Koo, Y.S. Kim, B.-I. Kim, K.M. Christensen, A comparative study of evacuation strategies for people with disabilities in high-rise building evacuation, Expert Syst. Appl. 40 (2) (2013) 408–417, https://doi.org/10.1016/j.eswa.2012.07.017.
[11] X. Zhao, R. Lovreglio, D. Nilsson, Modelling and interpreting pre-evacuation decision-making using machine learning, Autom. Constr. 113 (2020) 103140, https://doi.org/10.1016/j.autcon.2020.103140.
[12] S. Huang, J. Ji, Y. Wang, W. Li, Y. Zheng, A machine vision-based method for crowd density estimation and evacuation simulation, Saf. Sci. 167 (2023) 106285, https://doi.org/10.1016/j.ssci.2023.106285.
[13] J.C.P. Cheng, K. Chen, P.K.-Y. Wong, W. Chen, C.T. Li, Graph-based network generation and CCTV processing techniques for fire evacuation, Build. Res. Inf. 49 (2) (2021) 179–196, https://doi.org/10.1080/09613218.2020.1759397.
[14] Y. Ding, Y. Zhang, X. Huang, Intelligent emergency digital twin system for monitoring building fire evacuation, J. Build. Eng. 77 (2023) 107416, https://doi.org/10.1016/j.jobe.2023.107416.
[15] T.-Y. Lin, et al., Microsoft coco: common objects in context, in: Computer Vision–ECCV2014: 13th European Conference, Springer, Zurich, Switzerland, 2014, pp. 740–755. *September 6-12, 2014Proceedings, Part V 13*.

[16] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2010) 303–338.
[17] J. Cao, Y. Pang, J. Xie, F.S. Khan, L. Shao, From handcrafted to deep features for pedestrian detection: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2022) 4913–4934, https://doi.org/10.1109/TPAMI.2021.3076733.
[18] N. Chumuang, S. Hiranchan, M. Ketcham, W. Yimyam, P. Pramkeaw, T. Jensuttiwetchakult, Face detection system for public transport service based on scale-invariant feature transform, in: 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2020, pp. 1–6, https://doi.org/10.1109/iSAI-NLP51646.2020.9376819.
[19] B. Yang, W. Zhan, P. Wang, C. Chan, Y. Cai, N. Wang, Crossing or not? Context-based recognition of pedestrian crossing intention in the urban environment, IEEE Trans. Intell. Transp. Syst. 23 (6) (2022) 5338–5349, https://doi.org/10.1109/TITS.2021.3053031.
[20] H. Zhou, G. Yu, Research on pedestrian detection technology based on the SVM classifier trained by HOG and LTP features, Futur. Gener. Comput. Syst. 125 (2021) 604–615.
[21] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.
[22] P. Yang, G. Zhang, L. Wang, L. Xu, Q. Deng, M.-H. Yang, A part-aware multi-scale fully convolutional network for pedestrian detection, IEEE Trans. Intell. Transp. Syst. 22 (2) (2021) 1125–1137, https://doi.org/10.1109/TITS.2019.2963700.
[23] W. Liu, et al., Ssd: single shot multibox detector, in: European conference on computer vision, Springer, 2016, pp. 21–37.
[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," 2016. [Online]. Available: http://pjreddie.com/yolo/.
[25] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263–7271.
[26] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," *arXiv Prepr. arXiv1804.02767*, 2018.
[27] A. Bochkovskiy, C.-Y. Wang, and H.-Y.M. Liao, "Yolov4: optimal speed and accuracy of object detection," *arXiv Prepr. arXiv2004.10934*, 2020.
[28] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao, 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, 2023, pp. 7464–7475.
[29] P. Ultralytics, "YOLOv5," Github Repository. [Online]. Available: https://github.com/ultralytics/yolov5.
[30] P. Ultralytics, "YOLOv8," Github Repository. [Online]. Available: https://github.com/ultralytics/ultralytics.
[31] Z. Wang, Y. Ding, T. Zhang, X. Huang, Automatic real-time fire distance, size and power measurement driven by stereo camera and deep learning, Fire Saf. J. 140 (2023) 103891, https://doi.org/10.1016/j.firesaf.2023.103891.
[32] S. Li, L. Tong, C. Zhai, Extraction and modelling application of evacuation movement characteristic parameters in real earthquake evacuation video based on deep learning, Int. J. Disaster Risk Reduct. 80 (Oct. 2022) 103213, https://doi.org/10.1016/J.IJDRR.2022.103213.
[33] H. Ji, et al., Human abnormal behavior detection method based on T-TINY-YOLO, in: Proceedings of the 5th International Conference on Multimedia and Image Processing, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–5, https://doi.org/10.1145/3381271.3381273. ICMIP '20.
[34] H.H. Nguyen, T.N. Ta, N.C. Nguyen, V.T. Bui, H.M. Pham, D.M. Nguyen, YOLO based real-time human detection for smart video surveillance at the edge, in: 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), 2021, pp. 439–444, https://doi.org/10.1109/ICCE48956.2021.9352144.
[35] G. Overett, L. Petersson, N. Brewer, L. Andersson, N. Pettersson, A new pedestrian dataset for supervised learning, in: 2008 IEEE Intelligent Vehicles Symposium, 2008, pp. 373–378, https://doi.org/10.1109/IVS.2008.4621297.
[36] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: a benchmark, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1116–1124.
[37] C. Gu, et al., Ava: a video dataset of spatio-temporally localized atomic visual actions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6047–6056.
[38] D. Li, Z. Zhang, X. Chen, K. Huang, A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios, IEEE Trans. Image Process. 28 (4) (2018) 1575–1590.
[39] X. Liu, et al., Hydraplus-net: attentive deep features for pedestrian analysis, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 350–359.
[40] Y. Deng, P. Luo, C.C. Loy, X. Tang, Pedestrian attribute recognition at far distance, in: Proceedings of the 22nd ACM International Conference on Multimedia, 2014, pp. 789–792.
[41] S. Zhang, R. Benenson, B. Schiele, Citypersons: a diverse dataset for pedestrian detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3213–3221.
[42] D. Sharma, T. Hade, and Q. Tian, "Comparison Of Deep Object Detectors On A New Vulnerable Pedestrian Dataset," *arXiv Prepr. arXiv2212.06218*, 2022.
[43] W. Li, R. Zhao, X. Wang, Human reidentification with transferred metric learning, in: Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Springer, Daejeon, Korea, 2013, pp. 31–44. *November 5-9, 2012Revised Selected Papers, Part I 11*.
[44] D. Tzutalin, LabelImg, GitHub Repos 6 (2015).
[45] A. Paszke, et al., Pytorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019).
[46] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[47] Z. Zou, K. Chen, Z. Shi, Y. Guo, J. Ye, Object detection in 20 years: a survey, in: Proc. IEEE, 2023.
[48] N. Wojke, A. Bewley, D. Paulus, Simple online and realtime tracking with a deep association metric, in: 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3645–3649, https://doi.org/10.1109/ICIP.2017.8296962.
[49] A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Upcroft, Simple online and realtime tracking, in: 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 3464–3468, https://doi.org/10.1109/ICIP.2016.7533003.