DMU

Journal of Interdisciplinary Studies 8(2024) 1206-1217

Journal homepage: www.jims.dmu.edu.et

Volume 8(1), Jan 2024

DMU-JIDS

Real-Time Pedestrians Detection and Tracking in Disaster Areas Using Deep Learning Algorithms

Abebe Belay Adege[1,*]

[1]Department of Information Technology, College of Technology, Debre Markos University, Debre Markos, Ethiopia.

*Corresponding Author's Email: abbblybelay@gmail.com or Abebe_adege@dmu.edu.et

Abstract

Pedestrian detection and tracking in disaster-prone areas, such as valleys, deserts, beaches, and mountainous regions, present significant challenges due to the rugged terrain, dense vegetation, and the lack of advanced technology, especially in rural regions like Ethiopia. These challenging environments complicate the task of locating individuals, increasing the risk of them being lost during emergencies. The absence of reliable detection systems exacerbates the difficulty in providing timely rescue and relief operations. To address these challenges, we propose a deep learning-based model using video datasets collected through experimental methods, as well as data from publicly available sources. This combination of data enhances the model's ability to detect and track pedestrians effectively across a variety of disaster scenarios, ensuring its applicability to real-world conditions. Fourier Transform is applied to reduce noise and filter images, resulting in cleaner and more reliable inputs for the detection model. The processed datasets are then used within a transfer learning framework to develop a rapid and accurate pedestrian detection model based on the YOLOv7 architecture. This model is further integrated with Kalman filters to enable robust tracking of pedestrians, ensuring consistent performance even in dynamic and complex environments where traditional methods struggle. The proposed YOLOv7-based approach achieves a detection accuracy improvement of over 5% compared to existing state-of-the-art methods, including YOLOv2, ResNet, YOLOv3, YOLOv4, and YOLOv6. This improvement highlights the model's effectiveness in enhancing situational awareness and supporting emergency response efforts by accurately identifying and tracking individuals in disaster-affected areas. The conclusions from this research indicate that the proposed system provides a reliable and practical solution for aiding rescue and relief operations in challenging environments, ultimately contributing to improved disaster management outcomes.

Keywords: *Pedestrian detection and tracking; Transfer learning; Deep learning, YOLOv7.*

1. Introduction

In Ethiopia, the majority of the population resides in rural areas, where people often travel long distances to reach urban centers and exchange goods (Abate et al. 2020). These journeys are fraught with risks, as travelers frequently encounter natural disasters such as landslides, floods, and other hazards. Despite the dangers, the safety of these individuals is not adequately monitored by governmental or scientific means, due to the remote locations, insufficient infrastructure, and lack of attention to rural regions.

Traditional methods of tracking pedestrians in rural areas are often inadequate for surveillance in challenging environments such as mountainous, desert, or valley regions. Various solutions have been proposed by scholars to address these limitations, including sensor networks (Xu, Liang, and Xu 2014),Wi-Fi network base stations (X. Chen et al. 2023), and fixed cameras (J. Chen et al. 2024). However, these approaches are not cost-effective and require significant effort due to poor infrastructure. Datasets collected using these techniques in rural and disaster areas are significantly impacted by high data rate fluctuations and blockages caused by high-rise mountains or valleys. Moreover, challenges remain in identifying specific individuals in disaster locations, as pedestrian movements are dynamic across varying locations and altitudes. For instance, traditional methods struggle to track large areas with rapid changes due to their limited coverage and the mobility of individuals.

In contrast, UAV-based data collection techniques offer a more viable solution due to their portability, ease of access, and ability to capture data across diverse terrains, including hills and valleys (K. Wang et al. 2024). UAV-based data collection is employed to gather large datasets from sources such as disaster areas, where object detection and tracking applications are highly crucial. UAV technology provides a state-of-the-art solution for tracking systems, particularly in difficult-to-reach areas such as earthquake zones, floodplains, valleys, beaches, and hot regions (Hildmann and Kovacs 2019; Diaz et al. 2019). For example, in the Nile Gorge area in Ethiopia, as shown in Fig. 1, the terrain is vulnerable and difficult to monitor. While UAV technology equipped with high-quality cameras provides a straightforward solution for addressing these challenges, few have attempted to utilize it effectively for this purpose.

Figure 1. Sample Abay Gorge (Focus Area) Image

Various deep-learning algorithms have been proposed to address pediatric detection and tracking, including You Only Look Once (YOLOv3) (Zhong and Meng 2019), YOLOv2, DeepSort (Ma et al. 2019), and a hybrid approach combining YOLOv2, Long-Short Term Memory (LSTM), and Reinforcement Learning (RL) (C. X. Zhao et al. 2020). However, the tracking performances typically often struggle with accuracy in complex environments (Luo et al. 2018), particularly in locations with blurring effects due to mountainous terrain. Moreover, the majority of the work has not focused on rural and disaster areas. Moreover, although data optimization improves model accuracy, most studies apply object detection without it, resulting in less accurate outcomes compared to state-of-the-art performances (Zhong and Meng 2019).

In this research, a deep learning-based method for detecting and tracking multiple pedestrians is proposed, utilizing a transfer learning architecture. This approach combines experimental and publicly available datasets. The experimental data is collected using high-resolution images from smartphones and digital cameras, providing a comprehensive understanding of the phenomenon being studied. By integrating these datasets with publicly available sources, the approach offers a more complete perspective, allowing for deeper analysis and interpretation. This innovative combination leverages the strengths and adaptability of both types of datasets, leading to more accurate and reliable conclusions. YOLOv7 with transfer learning is employed for robust object detection, allowing the newly developed model to leverage knowledge from existing models and efficiently generate a more accurate and adequate new model. In the new model, the Kalman filter is integrated with YOLOv7 to associate frames and track pedestrians in motion. The following are the main contributions of this work:

1. By integrating experimental and publicly available data, the approach provides a more comprehensive understanding of the phenomenon, facilitating richer analysis and more informed conclusions.
2. A fast and accurate pedestrian detection model is developed using YOLOv7, incorporating Fourier Transform and transfer learning to minimize noise and improve data quality.
3. YOLOv7 is combined with Kalman filters for robust tracking, enabling the association of sequential frames and enhancing target tracking performance.
4. The proposed system outperforms state-of-the-art methods (YOLOv2, ResNet, YOLOv3, YOLOv4, and YOLOv6) by achieving higher detection accuracy through YOLOv7 with transfer learning.

The remaining sections are structured as follows: Section II reviews relevant surveys. Section III outlines the proposed method and algorithm usage. Section IV details simulation setups, and findings and provides an analysis. Section V concludes the paper and discusses future directions.

2. Related Works

Pedestrian detection and tracking in disaster areas is a challenging task that requires efficient and accurate methods to ensure timely response and rescue efforts. Recently, deep learning algorithms have shown promising results in various computer vision applications, including object detection and tracking. Object detection and tracking are crucial tasks in various applications, including object recognition and classification. Tracking involves the process of associating the location of a targeted object from one frame to the next in a sequence of images or visual data (Bohush et al. 2020). Multi-object tracking tasks comprise the association or integration of segmented motion objects in continuous images of a video sequence. Several studies have been conducted on object detection and tracking using deep learning-based approaches. For instance, (Ye et al. 2018) proposed a deep learning-based model to estimate UAV motion, utilizing CNN for feature extraction and the Kalman filter to improve temporal consistency and reduce false alarm rates. The precision ranged from 81% to 82%, and accuracy reached 95%. However, the multi-stage detection technique used in this approach is time-consuming.

In contrast, YOLOv3 (L. Zhao and Li 2020) showed better computational costs compared to other algorithms such as Single Shot MultiBox Detector (SSD), Deconvolutional Single Shot Detector (DSSD), Region-based Fully Convolutional Networks (R-FCN), and Residual Network (ResNets). Most region-based algorithms, including R-CNN and fast R-CNN, face longer and more complex calculations to extract features, which leads to slower computational costs and loss of relevant image information during continuous tracking. YOLOv1 and YOLOv2 are sensitive to image backgrounds and positions, resulting in unbalanced detection speed and accuracy (Diaz et al. 2019) and (Dutra and Orth, 2020.). In Dutra and Orth, 2020), the authors aimed to design a system using a hybrid of YOLOv2 and deep sorting, with the Kalman filter used for backend error smoothing during tracking. The system performance showed an average accuracy ranging from 68.7% to 86.8%.

(Behera et al. 2017) used CNN algorithms in transfer learning for Crowd Density Detection and Classification on video surveillance systems, achieving an average accuracy of 96.66%. In (Narmadha et al. 2023), a ResNet-based Faster R-CNN was applied to detect and track objects, with manual data annotation and preprocessing followed by evaluation on three benchmark video datasets. In contrast to these approaches, (L. Zhao and Li 2020) used the SSD technique without preprocessing, while (Bochkovskiy, Wang, and Liao 2020) utilized the YOLO approach for object detection, both eliminating the need for additional region proposal networks.

Several studies have focused on pedestrian detection using deep learning algorithms. For example, (Li et al. 2020) proposed a real-time pedestrian detection system using a DNN and a kalman filter for tracking. The system achieved a precision of 92% and a recall of 95%. In terms of tracking, (Csönde, Sekimoto, and Kashiyama 2022) proposed a real-time pedestrian tracking system using a combination of CNN and optical flow estimation. The system tracked pedestrians with an average precision of 90% and an average recall of 92%. Another study (Sighencea, Stanciu, and Căleanu 2021) discussed a recurrent neural network (RNN) to track pedestrians in videos, achieving an average precision of 95% and an average recall of 98%. In (López-Randulfe et al. 2021), the authors proposed a system that uses a CNN to detect pedestrians in images taken from aerial vehicles during disasters. The system achieved an accuracy of 95% and a precision of 90%. Another study (Y. Wang, Shi, and Wu 2019) used a DNN to track pedestrians in videos taken from ground-based cameras during disasters, achieving an average precision of 92% and an average recall of 95%.

Other studies have focused on improving the performance of pedestrian detection and tracking systems by combining multiple techniques. For example, (Hildmann and Kovacs 2019; Diaz et al. 2019) proposed a system that uses a CNN to detect pedestrians and a Kalman filter to track them. The system achieved an accuracy of 98% and a precision of 95%. Another study used an RNN to track pedestrians and a CNN to detect them, achieving an average precision of 95% and an average recall of 98%.

Overall, these studies demonstrate the potential of deep learning algorithms for pedestrian detection and tracking in disaster areas. However, there is stillroom for improvement, especially in terms of real-time performance and robustness to varying environmental conditions.

3. Proposed Methods

3.1. Proposed Architecture

Fig. 2 shows the architecture of a pedestrian detection and tracking system, combining YOLOv7 with a Kalman Filter. The system uses experimental data from real-world environments like the Abay Gorge region, along with publicly available datasets, to train and validate the model. Then, preprocessing steps such as image normalization, resizing, and noise reduction are applied. The data is then split into training and testing sets. YOLOv7 is fine-tuned during training, resulting in a model capable of accurate pedestrian detection, which was saved as a YOLOv7 (.h5) file. For tracking, the system employs a Kalman Filter, which predicts and updates the positions of detected pedestrians, ensuring precise monitoring in dynamic environments. The combined use of YOLOv7 for detection and the Kalman Filter for tracking enables real-time, accurate surveillance. The testing set is used to validate the system, ensuring it performs well in real-world scenarios. This system is ideal for applications such as surveillance, disaster management, and crowd control, where real-time detection and tracking are critical.

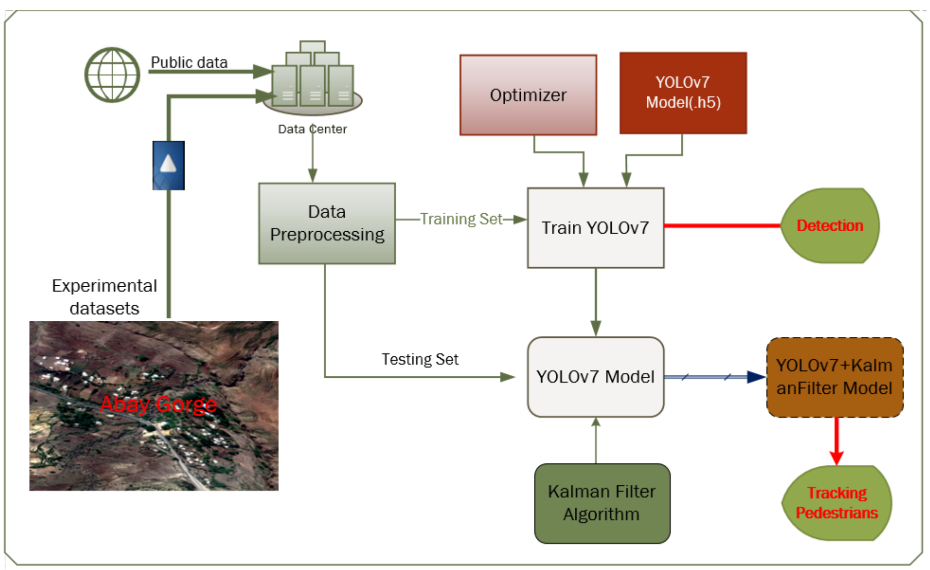

Figure 2. The architecture of the proposed system.

3.2. Data sources

This study utilized two types of data sources. The first set of datasets was collected through fieldwork using Nikon cameras and Samsung M12 and M13 smartphones. A total of 50 video files, each 10 seconds long, were captured in the Abay Valley region of Ethiopia, Amhara region. Data was gathered from three distinct locations, each offering different perspectives for easily capturing pedestrians. The second source consisted of 50 video files, also 10 seconds long, focused on rural communities in sloping areas. These videos were sourced from platforms like YouTube (e.g., (https://www.youtube.com/watch?v=IJCy64adY3Y and https://www.youtube.com/watch?v=NZ55BsC2kn8). All video files from both sources were stored in MP4 format and later converted to JPG format for preprocessing. The data collection period spanned from January to June 2023.

3.3.Preprocessing

Due to challenges posed by low image quality—such as dust and reflections from disaster-prone areas—additional preprocessing steps were applied, including histogram equalization, and further noise removal to enhance the overall quality of the datasets. Image noise is removed using Gaussian filtering, which reduces noise and helps minimize overfitting problems in model development (Zhong and Meng 2019), as illustrated in Equation (1).

$$I_{\text{filtered}}(x,y) = \frac{1}{2\pi\sigma 2} \exp\left(-\frac{x^2+y^2}{2\sigma 2}\right) * I(x,y) \quad \ldots\ldots(1)$$

Here, $I_{filtered}\ (x,y)$ represents a Gaussian blur applied to reduce noise in the image, $2\sigma_2$ is the standard deviation of the Gaussian distribution, x and y are pixel coordinates, and "$*$" denotes convolution. To ensure uniformity in image sizes, we resized them to a standard dimension of 224x224 RGB. This adjustment was necessary as the images were captured at varying distances from the camera. To enhance image quality, we employed histogram equalization, detailed in Equation 2, which adjusts image intensities to improve overall contrast.

$$N\,(P(x\ y)) = \text{round}\left(\frac{cdf(p(x\ y)) - cdfmin}{(Rx \times Cx) - cdfmin} \times (L-1)\right) \quad \ldots\ldots(2)$$

Where, $N\,(P(x\ y))$ is histogram equalization, cdf refers to cumulative frequency, Cdf_{min} is a minimum value of cumulative distribution function, $Cdf(p(x,y))$ is an intensity of the current pixel, Rx & Cx are product of number of pixels in rows and columns and L is number of intensities.

The segmentation process in YOLOv7 enhances object detection by dividing the input image into a grid of cells, where each cell predicts multiple bounding boxes with confidence scores. These scores indicate both the likelihood of an object and the accuracy of the bounding box location (López-Randulfe et al. 2021). Each cell also predicts class probabilities, allowing YOLOv7 to classify objects like pedestrians or vehicles (C.-Y. Wang, Bochkovskiy, and Liao 2022). YOLOv7 introduces anchor boxes, predefined bounding boxes of various sizes and aspect ratios, to improve detection accuracy for objects of different shapes. Additionally, Non-Maximum Suppression (NMS) is applied to eliminate redundant bounding boxes, ensuring each object is detected only once by selecting the one with the highest confidence score. This builds on prior methods in adaptive object tracking and vehicle tracking using Kalman filters and Deep Sort with low-confidence track filtering. The Fourier Transform exponential function for image enhancement can be defined as shown in Equation (3).

$$F(U) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \varepsilon}\, d \quad \ldots\ldots(3)$$

Where $f(x)$ is a function, $e^{-2\pi i x \varepsilon} d$ is the Fourier Transform exponential function, $F(U)$. Then, we apply confidence of the prediction that is used to detect objects in a bounding box, which is calculated as indicated in Equation (4), which is an interactive class-based object detection technique (Lee et al. 2022).

$$Pr(Pit, ft \mid V, CN) = Pr(Pit \mid ft, Ci) * Pr(ft \mid Ci) * Pr(Ci) \quad \ldots\ldots(4)$$

Where i = [1,2,⋯,k] denotes the i^{th} person at a time t, *it* is the frame of time t, C_N is N

numbers of classes, and P_i^t is the person that could be detected from a given video (V).

The effectiveness of the model was measured against various ground-truth values during the in-person class, selecting the optimal fit based on bounding box (Bbox) data. The class probability, CN, varies, and those probabilities that meet a defined threshold are selected as the ground truth. Equation (5) demonstrates the computation of the Intersection over Union (IoU):

$$IoU[ID = i] = \frac{area(C_i) \cap area(G)}{area(C_i) \cup area(G)} \quad \text{……………………………………………………(5)}$$

Where C_i is the i^{th} Bbox in a frame and G is the Bbox of the ground truth. Then, the loss function of IoU is formulated as indicated in Equation (6).

$$L_{IoU} = 1 - IoU + \frac{p^2(b, b^{gt})}{C^2} + av \quad \text{………………………………………………………………(6)}$$

Where α denotes the hyperparameter controlling the trade-off and is set $a = \frac{v}{((1-IoU)+v)}$ according to previous literature. To associate sequences of Bboxes in different frames, we apply the Kalman filter (Ait Abdelali et al. 2016. Let A denote the processing matrix each time $t > 0$, then the state prediction of a Kalman filter is shown in Equation (7):

$$St = A \times S_t - 1 \quad \text{…………………………………… (7)}$$

Where S_t and S_{t-1} restate predictions of the Kalman filter, and the vector of the process state at the time $t - 1$, respectively.

Once a set of satisfactory values is found, the most similar value pair is calculated using the cosine similarity (Kumar et al. 2023), by estimating distances between vectors. This technique is more straightforward than the Mahalanobis approach since it is not relatively affected easily by the size of images compared to the Mahalanobis technique. The cosine similarity, sim (F_c, F_n), between vectors of the successive frames is calculated as shown in Equation (8):

$$sim(F_c, F_n) = \frac{\sum_{i=1}^{k} F_{c,i} F_{n,k}}{\sqrt{\sum_{i=1}^{k} F_{c,i}^2 \sum_{i=1}^{k} F_{n,k}^2}}, \quad \text{…………………………………………………………(8)}$$

Where F_c and F_n represent features of the bounding box in the current and the next frames, respectively. $F_{c,i}$ is the i^{th} element of F_c, and $F_{n,k}$ is the k^{th} element of the next frame. After the best similar Bboxes are obtained using cosine similarity, the logistic regression is applied to fall the prediction ranges between [0, 1] to satisfy the conditional probability. In this study, we utilize multiple criteria to assess the effectiveness of the proposed models. These criteria encompass accuracy, precision, and recall.

4. Simulation Setup and Results

4.1. Simulation Setup

In our experiments, we varied the heights and locations from which video data were captured at disaster sites to assess how environmental conditions at different altitudes and distances affect the performance of our algorithms. Each experiment involved extensive image preprocessing, data augmentation, image segmentation, feature extraction, and classification using the Adam optimizer. The Adam optimizer was selected for its adaptive learning rate and effective regularization methods, which help mitigate overfitting. Testing was conducted with different epoch sizes and a fixed batch size of 32, with the optimal epoch automatically determined during training. We utilized data splits of 70%:30%, 80%:20%, and 85%:15% for training and testing, as shown in Table 1, with optimal performance observed at the

85%:15% ratio. This work faced significant challenges, including integrating multiple algorithms into a cohesive model and optimizing the quality of data collected from disaster areas during fieldwork.

Table 1. YOLOv7 performances at different ratios.

Training-Testing Ratios	Average Accuracy	Optimum Performances
70%:30%	93.17%	85%:15% ratio
80%:20%	96.05%	
85%:15%	97.53 %	

The performance of the YOLOv7 model was evaluated at different training-testing ratios, as shown in Table 1. With a 70% training and 30% testing split, the model achieved an accuracy of 93.17%. Increasing the training data to 80% and reducing the testing data to 20% improved accuracy to 96.05%. The highest accuracy, 97.53%, was obtained when the training data comprised 85% of the dataset, with the testing data at 15%. These results suggest that the model's performance improves as the proportion of training data increases, with optimal performance observed at the 85%:15% training-testing ratio.

4.2. Results and Discussions

Figure 3 illustrates the process of selecting the optimal IoU value, crucial for accurate object detection and tracking. Selecting an inappropriate IoU threshold can result in misclassifications and missed targets, thereby affecting localization accuracy. To ensure unbiased selection and robust conclusions, we compare system performance across IoU thresholds of 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, and 0.85. Through this comparative analysis, an IoU threshold of 0.7 is identified as yielding the highest precision, as depicted in Figure 3. These findings underscore the significant impact of IoU values on model performance, highlighting the need for careful consideration and optimization in object detection tasks.

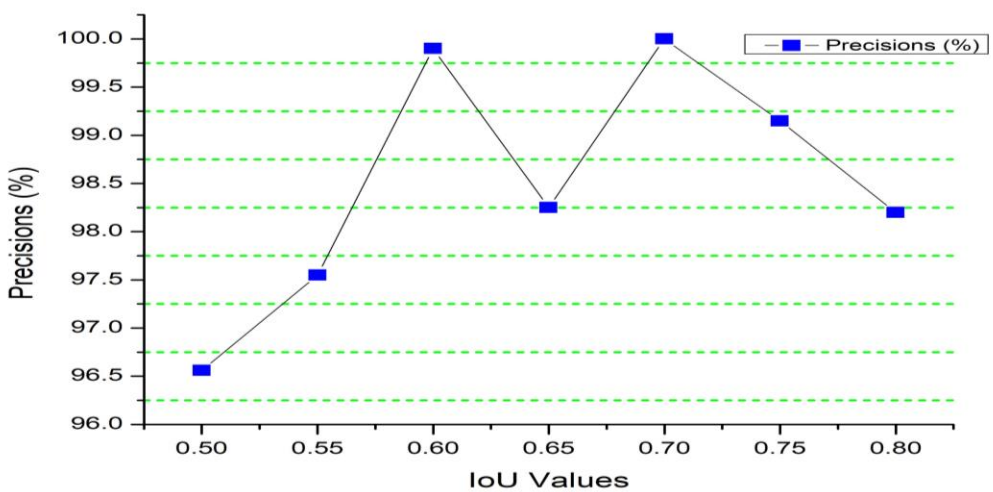

Figure 3. Individual detection performances of the proposed model in different IoU values.

Table 2 presents the detection and tracking performance of individuals using YOLOv7 alone and in combination with Kalman filter algorithms, respectively. Detection performance is assessed through precision and accuracy metrics while tracking performance is evaluated using confidence scores. In all activities, there is a significant overlap between estimated and ground-truth bounding boxes, resulting in accurate detection of persons in motion. This leads to comparable average detection and tracking performance. Specifically, for different experimental locations—location 1, location 2, and location 3—precision rates of 98.7%, 97.9%, and 98.33% are achieved, respectively. Consequently, the average precision of the proposed model stands at 98.31%, the average accuracy is 97,53 and the average confidence score is 97.61, respectively.

Despite there is a presence of significant noise in the experimental image datasets, we achieved accurate and robust detection and tracking performance due to the superior capabilities of YOLOv7 and the continuous application of optimization techniques such as early stopping and dropout. However, variations in camera locations led to minor discrepancies. For instance, wood and small objects were occasionally misclassified as humans, and some individuals were not correctly grouped into in-person classes. Nevertheless, the accuracy and average precision of YOLOv7 remained above 95%. The confidence scores for the combination of YOLOv7 and the Kalman filter across different testing locations exceeded 96%, demonstrating the robustness of the proposed model. Generally, the accuracy was consistent across varying heights. Our proposed system, which integrates YOLOv7 and Kalman filters, effectively addresses these challenges due to YOLOv7's rapid detection capability and the Kalman filter's bounding box association capacity. Additionally, optimization technique es such as IoU and cosine similarity were employed to further enhance accuracy.

Table 2. YOLOv7 -based Pedestrian Detection and Tracking Performances.

Detection Performances			
Scenarios	**Precision**	**Accuracy**	**Confidence- Score**
Location1	98.7 %,	98.38 %	96.25 %
Location2	97.9 %	96.30 %	99.23 %
Location3	98.33 %	97.9 %	97.36 %
Average values	98.31 %	97.53 %	97.61 %

Table 3 presents a comparison of various state-of-the-art algorithms [YOLOv2, ResNet, YOLOv3, YOLOv4] with the proposed model, YOLOv7. YOLOv2 shows the lowest performance, with a precision of 55.72% and an accuracy of 52.34%. These low numbers suggest that YOLOv2 struggles with both identifying objects correctly and maintaining overall accuracy. ResNet performs better, with a precision of 79.1% and an accuracy of 76.90%. Although this is an improvement, ResNet still falls short of the more recent versions of YOLO in both precision and accuracy. YOLOv3 enhances detection capabilities with 91% precision and 90.42% accuracy, and YOLOv4 achieves 94.8% precision and 91.7% accuracy. However, YOLOv7 outperforms all previous algorithms, attaining the highest precision of 98.31% and accuracy of 97.53%. This comparison highlights the superior performance of YOLOv7 in detecting and tracking pedestrians, demonstrating significant improvements in both precision and accuracy over other state-of-the-art algorithms.

Table 3. Comparisons of YOLOv7 with other algorithms.

Comparison of Detection and Tracking (Testing)		
Algorithms	**Precision**	**Accuracy**
YOLOv2	55.72 %	52.34 %
ResNet	79.1 %	76.90 %
YOLOv3	91 %	90.42 %
YOLOv4	94.8 %	91.7 %
OURS(YOLOv7)	98.31 %	97.53 %

Table 4 provides a comparative analysis of tracking performance between the YOLOv7-Kalman Filter and YOLOv6-Kalman Filter models across three different types of datasets: publicly available data, experimental data, and hybrid datasets (a combination of both publicly available and experimental data). The YOLOv7-Kalman filter consistently outperforms the YOLOv6-Kalman filter across all scenarios. For publicly available data, the YOLOv7-Kalman filter achieves an accuracy of 96.76% and a precision of 97.98%, compared to the YOLOv6-Kalman filter's 87.00% accuracy and 85.10% precision. In the experimentally collected data, YOLOv7 maintains its superior performance with an accuracy of 96.11% and precision of 98.00%, while YOLOv6 records a lower accuracy of 85.12% and precision of 81.80%.

When evaluating hybrid datasets, which combine both publicly available and experimental data, the YOLOv7-Kalman filter achieves the highest performance with an accuracy of 97.53% and a precision of 98.31%. In contrast, the YOLOv6-Kalman filter shows a significant decrease in performance, with accuracy dropping to 79.18% and precision at 85.20%. The superior performance of the YOLOv7-Kalman filter can be attributed to the combination of YOLOv7's advanced detection capabilities and the Kalman filter's robust tracking efficiency. Continuous optimization techniques, along with YOLOv7's fast and precise detection capabilities, further enhance its performance. This makes the YOLOv7-Kalman filter a more reliable choice for accurate and precise pedestrian detection and tracking across various data scenarios.

Table 4. YOLOv7-Kalman filter tracking performances.

Data Sources	Accuracy		Precision	
	YOLOV7-Kalman Filter (%)	YOLOV6-Kalman Filter (%)	YOLOV7-Kalman Filter (%)	YOLOV6-Kalman Filter (%)
Publicly Available Data Only	96.76%	87.00%	97.98%	85.10%
Experimental Data Only	96.11%	85.12%	98.00%	81.80%
Hybrid Datasets	97.53 %	79.18%	98.31 %	85.20%

5. Conclusion

This research presents a deep learning-based model for real-time pedestrian detection and tracking in disaster-prone areas, utilizing both experimental and publicly available datasets. The use of Fourier Transform for noise reduction, along with transfer learning, enhances the model's ability to detect pedestrians effectively across challenging terrains like valleys, deserts, and mountainous regions. The model's integration of YOLOv7 for precise detection and Kalman filters for robust tracking enables it to maintain high performance, with detection accuracy reaching up to 97.53% and precision of 98.31%. This performance surpasses existing state-of-the-art algorithms such as YOLOv2, ResNet, YOLOv3, YOLOv4, and YOLOv7 as demonstrated in multiple testing scenarios. The application of optimization techniques, such as early stopping and IoU threshold optimization, further strengthens the reliability of the system, even in the presence of noise and varying environmental conditions. The findings of this study highlight the potential of the YOLOv7-based model to contribute meaningfully to disaster management by enhancing situational awareness and supporting rescue and relief operations. Additionally, its adaptability makes it useful in other sectors, including security, military operations, and agriculture, where rapid and accurate detection and tracking are essential. Future research will focus on incorporating zero-delay image processing and mobile data fusion to improve crowd-sensing capabilities in urban environments, offering even broader applications for real-time detection and monitoring technologies using drone technology.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgment

This research is not supported by any financial sources.

References

Abate, Gashaw T., Mekdim Dereje, Kalle Hirvonen, and Bart Minten. 2020. “Geography of Public Service Delivery in Rural Ethiopia.” *World Development* 136 (December). https://doi.org/10.1016/j.worlddev.2020.105133.

Ait Abdelali, Hamd, Fedwa Essannouni, Leila Essannouni, and Driss Aboutajdine. 2016. “An Adaptive Object Tracking Using Kalman Filter and Probability Product Kernel.” *Modelling and Simulation in Engineering* 2016 (February). https://doi.org/10.1155/2016/2592368.

Behera, Rabi Narayan, Kajaree Das, B Tech, and Asst Professor. 2017. “A Survey on Machine Learning: Concept, Algorithms and Applications International Journal of Innovative Research in Computer and Communication Engineering A Survey on Machine Learning: Concept, Algorithms and Applications.” *Article in International Journal of Innovative Research in Computer and Communication Engineering*. https://doi.org/10.15680/IJIRCCE.2017.

Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. “YOLOv4: Optimal Speed and Accuracy of Object Detection,” April. http://arxiv.org/abs/2004.10934.

Bohush, Rykhard, Sergey Ablameyko, Sviatlana Ihnatsyeva, and Yahor Adamovskiy. 2020. “Object Detection Algorithm for High Resolution Images Based on Convolutional Neural Network and Multiscale Processing.”

Chen, Jianguo, Kenli Li, Qingying Deng, Keqin Li, and Philip S. Yu. 2024. “Distributed Deep Learning Model for Intelligent Video Surveillance Systems with Edge Computing.” *IEEE Transactions on Industrial Informatics*. https://doi.org/10.1109/TII.2019.2909473.

Chen, Xiaoqian, Kang Chen, Minxiao Wang, and Ruopu Li. 2023. “Roles of Wireless Networks in Bridging the Rural Smart Infrastructural Divide.” *Infrastructures* 8 (11).https://doi.org/10.3390/infrastructures8110159.

Csönde, Gergely, Yoshihide Sekimoto, and Takehiro Kashiyama. 2022. “Online Real-Time Pedestrian Tracking from Medium Altitude Aerial Footage with Camera Motion Cancellation.” *Computer Vision and Image Understanding* 217 (March). https://doi.org/10.1016/j.cviu.2022.103386.

Diaz, Jade Luke, Giesther Gapasangra, Paulo Daniel Carandang, and Mary Jane C. Samonte. 2019. “Local Disaster Risk Reduction Management System Utilizing Unmanned Aerial Vehicles.” In *ACM International Conference Proceeding Series*, 24–28. Association for Computing Machinery. https://doi.org/10.1145/3374549.3374581.

Dutra, Emanuel, and Rene Orth. n.d. “Robustness of Process-Based versus Data-Driven Modeling in Changing Climatic Conditions.” https://doi.org/10.1175/JHM-D.

Hildmann, Hanno, and Ernö Kovacs. 2019. “Review: Using Unmanned Aerial Vehicles (Uavs) as Mobile Sensing Platforms (Msps) for Disaster Response, Civil Security and Public Safety.” *Drones*. MDPI AG. https://doi.org/10.3390/drones3030059.

Kumar, Sunil, Sushil Kumar Singh, Sudeep Varshney, Saurabh Singh, Prashant Kumar, Bong Gyu Kim, and In Ho Ra. 2023. “Fusion of Deep Sort and Yolov5 for Effective Vehicle Detection and Tracking Scheme in Real-Time Traffic Management Sustainable System.” *Sustainability (Switzerland)* 15 (24). https://doi.org/10.3390/su152416869.

Lee, Chunggi, Seonwook Park, Heon Song, Jeongun Ryu, Sanghoon Kim, Haejoon Kim, Sérgio Pereira, and Donggeun Yoo. 2022. “Interactive Multi-Class Tiny-Object Detection.” https://github.

Li, Tao, Yitao Ma, Hui Shen, and Tetsuo Endoh. 2020. “FPGA Implementation of Real-Time Pedestrian Detection Using Normalization-Based Validation of Adaptive Features Clustering.” *IEEE Transactions on Vehicular Technology* 69 (9): 9330–41. https://doi.org/10.1109/TVT.2020.2976958.

López-Randulfe, Javier, Tobias Duswald, Zhenshan Bing, and Alois Knoll. 2021. “Spiking Neural Network for Fourier Transform and Object Detection for Automotive Radar.” *Frontiers in Neurorobotics* 15 (June). https://doi.org/10.3389/fnbot.2021.688344.

Luo, Wenhan, Peng Sun, Fangwei Zhong, Wei Liu, Tong Zhang, and Yizhou Wang. 2018. “End-to-End Active Object Tracking via Reinforcement Learning.”

Ma, Xinpei, Junye Cheng, Liubing Dong, Wenbao Liu, Jian Mou, Ling Zhao, Jinjie Wang, et al. 2019. “Multivalent Ion Storage towards High-Performance Aqueous Zinc-Ion Hybrid Supercapacitors.” *Energy Storage Materials* 20 (July):335–42. https://doi.org/10.1016/j.ensm.2018.10.020.

Narmadha, C., T. Kavitha, R. Poonguzhali, V. Hamsadhwani, Ranjan Walia, Monia, and B. Jegajothi. 2023. "Robust Deep Transfer Learning Based Object Detection and Tracking Approach." *Intelligent Automation and Soft Computing* 35 (3): 3613–26. https://doi.org/10.32604/iasc.2023.029323.

Sighencea, Bogdan Ilie, Rareș Ion Stanciu, and Cătălin Daniel Căleanu. 2021. "A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction." *Sensors*. MDPI. https://doi.org/10.3390/s21227543.

Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2022. "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," July. http://arxiv.org/abs/2207.02696.

Wang, Kaiwen, Lammmert Kooistra, Ruoxi Pan, Wensheng Wang, and João Valente. 2024. "UAV-Based Simultaneous Localization and Mapping in Outdoor Environments: A Systematic Scoping Review." *Journal of Field Robotics* 41 (5): 1617–42. https://doi.org/10.1002/rob.22325.

Wang, Yong, Wei Shi, and Shandong Wu. 2019. "Robust UAV-Based Tracking Using Hybrid Classifiers." *Machine Vision and Applications* 30 (1): 125–37. https://doi.org/10.1007/s00138-018-0981-4.

Xu, Xu, Weifa Liang, and Zichuan Xu. 2014. "Remote Monitoring Cost Minimization for an Unreliable Sensor Network with Guaranteed Network Throughput." *Information Processing in Agriculture* 1 (2): 83–94. https://doi.org/10.1016/j.inpa.2014.10.001.

Ye, Dong Hye, Jing Li, Qiulin Chen, Juan Wachs, and Charles Bouman. 2018. "Deep Learning for Moving Object Detection and Tracking from a Single Camera in Unmanned Aerial Vehicles (Uavs)." In *IS and T International Symposium on Electronic Imaging Science and Technology*. Society for Imaging Science and Technology. https://doi.org/10.2352/ISSN.2470-1173.2018.10.IMAWM-466.

Zhao, Chang Xin, Jia Ning Liu, Bo Quan Li, Ding Ren, Xiao Chen, Jia Yu, and Qiang Zhang. 2020. "Multiscale Construction of Bifunctional Electrocatalysts for Long-Lifespan Rechargeable Zinc–Air Batteries." *Advanced Functional Materials* 30 (36). https://doi.org/10.1002/adfm.202003619.

Zhao, Liquan, and Shuaiyang Li. 2020. "Object Detection Algorithm Based on Improved YOLOv3." *Electronics (Switzerland)* 9 (3). https://doi.org/10.3390/electronics9030537.

Zhong, Mingliang, and Fei Meng. 2019. "A YOLOv3-Based Non-Helmet-Use Detection for Seafarer Safety Aboard Merchant Ships." In *Journal of Physics: Conference Series*. Vol. 1325. Institute of Physics Publishing. https://doi.org/10.1088/1742-6596/1325/1/012096.