

A novel YOLOv8 architecture for human activity recognition of occluded pedestrians

Shaamili Rajakumar, Ruhan Bevi Azad

Department of Electronics and Communication Engineering, College of Engineering and Technology,
SRM Institute of Science and Technology, Kattankulathur Campus, Chengalpattu, Tamil Nadu, India

Article Info

Article history:

Received Apr 17, 2024

Revised Jul 10, 2024

Accepted Jul 17, 2024

Keywords:

Adverse weather conditions
Adaptive spatial feature fusion
Bidirectional feature pyramid
network
C2f module
Human activity recognition
Real time videos/images
You look only once v8

ABSTRACT

Perception is difficult in video surveillance applications because of the presence of dynamic objects and constant environmental changes. This problem worsens when bad weather, including snow, rain, fog, dark nights, and bright daylight, interferes with the quality of perception. The proposed work aims to enhance the accuracy of camera-based perception for human activity detection in video surveillance during adverse weather conditions. To identify primary human activities, including walking on the road during severe weather, transfer learning from many adverse conditions using real-time images or videos has been proposed as an improvement for you look only once v8 (YOLOv8)-based human activity recognition in poor weather conditions. We collected and sorted training rates into frames from videos depicting human walking activity, their combined forms, and other subgroups, such as running and standing, based on their characteristics. The assessment of the detection efficiency of the previously described images and subgroups led to a comparison of the training weights. The use of real-time activity images for training greatly enhanced the detection performance when comparing the proposed test results to the existing YOLO base weights. Furthermore, a notable improvement in human activity efficiency was obtained by utilizing extra images and feature-related combinations of data techniques.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Ruhan Bevi Azad

Department of Electronics and Communication Engineering, College of Engineering and Technology,

SRM Institute of Science and Technology, Kattankulathur Campus

Chengalpattu, Tamil Nadu, India

Email: ruhanb@srmist.edu.in

1. INTRODUCTION

Automated people detection is essential, with numerous useful applications in transportation management, security systems, video surveillance, and crowd control. This broad range of applicability is categorized into human identification and human verification. While human identification aims to locate and identify a person among a set of various items, the primary focus is on identifying whether a human exists among a set of different components in an image or video. Despite extensive research, many real-world applications still view human identification as a difficult task. Human stances are flexible and help to differentiate between groups of individuals based on body type, size, and shape. Additionally, noise, occlusions, and the surrounding environment have a significant impact on the performance of human identification and tracking in real time.

In recent decades, computer vision for human identification has widely employed deep learning-based techniques. Convolutional neural networks (CNN) have become the most widely used technique for

solving real-time challenges. Various robust and selective CNN topologies are currently used to solve problems related to object detection, crowd counting, and passenger flow computation. A subset of artificial intelligence known as neural networks is capable of efficiently learning features from vast human datasets. CNN is now known to be the best option for identifying people in crowded environments. Since the recent introduction of convolutional neural networks, the most controversial and important advancements in human recognition have remained unclear. Using the collected features, convolution and pooling procedures in CNNs are able to identify and recognize humans in obscured situations and unrestricted environments. However, putting context awareness and personalization into practice requires gathering, analyzing, and storing enormous volumes of personal data, which raises privacy issues. Safeguarding confidential data and upholding user confidence are critical as smart living solutions and apps become more integrated into users' daily lives. Therefore, it is crucial to strike a balance between protecting privacy and using data for personalization. Advanced privacy-preserving methods [1] such as encryption and anonymization must guarantee that user data are kept private to reach this balance. Finally, data accessibility is critical to human action recognition (HAR) efficient operation in smart home apps and services. These systems' capacity to function depends on the availability and consistency of data since they need it constantly to make judgments and provide individualized experiences. The dynamic nature of smart living environments necessitates the maintenance of data consistency across multiple platforms and devices, making data accessibility particularly challenging. Creating strong infrastructure and data management plans is essential for HAR implementation in smart living.

2. RELATED WORK

Haq *et al.* [2] created a deep learning algorithm for trustworthy human identification and tracking in an obstructed environment. Combining several data augmentation approaches has created complex settings for the human look. Only the use of a fused loss function, which combined the SoftMax and localization losses, increased the dependability and efficiency of the deep model. The results of the study using the INRIA human dataset and pedestrian parsing for monitoring show that the proposed method makes it easier to recognize and keep an eye on people in noisy and obscured environments. The most recent versions of you only look once v8 (YOLOv8) and an enhanced YOLOv5s model were compared to construct a computer vision-based photovoltaics (PV) fault detection system [3]. Compared with the baseline version, the enhanced version of YOLOv5s was able to detect defects in PV modules with a decent balance between performance and efficiency. To enhance network feature extraction, improved YOLOv5 adds the global attention mechanism (GAM) unit to the neck and backbone after starting with YOLOv5s as the baseline. To optimize the network's lower-level feature spatial data and upper-level semantic information, the adaptive spatial feature fusion (ASFF) was also added to the head branch. Ultimately, DIOU was used in place of the original non-maximum suppression (NMS) loss function.

Some researchers have used the bidirectional feature pyramid network (BiFPN) method [4], which replaces the path aggregation network (PANET) structure. This made the model increasingly smaller and better at fusing features. The SimSPPF module improves the efficiency of the spatial pyramid pooling phase to expedite the model's detection process. The author introduces the LSK attention strategy, which has dynamic large convolutional kernels to improve the accuracy of the model. The outcome of the investigation demonstrated how well the improved model can identify road faults in images taken by drones and cameras installed on cars. Novel techniques for feature fusion [5] and network architectures were introduced. The network's learning ability was significantly enhanced. The experiment involved conducting tests and comparisons on three datasets: the Visdrone, the Tiny Person, and the PASCAL VOC2007 datasets. A thorough study and several trials proved the viability of each improvement component. The performance of DC-YOLOv8 exceeded that of the other detectors in terms of speed and accuracy. Capturing small targets in diverse and intricate situations was more manageable. To recognize objects in aerial scenes using unmanned aerial vehicles (UAVs), Wang *et al.* [6] introduced a model named UAV-YOLOv8. This model is based on YOLOv8 and aims to improve identification efficiency while taking into account platform resources. Initially, an adaptive sample distribution method is incorporated into the WIoU v3 loss function, resulting in a notable decrease in the model's focus on exceptional samples and an enhancement in overall performance. Furthermore, the backbone network incorporates BiFormer, an advantageous dynamic sparse attention mechanism. This enhances the model's ability to concentrate on crucial information within the feature maps, resulting in improved detection performance.

The LAR-YOLOv8 model [7] was applied to three datasets: the NWPU VHR-10, RSOD, and CARPK datasets. The results show that the proposed LAR-YOLOv8 model achieves improvements in the mean average precision (mAP) of 4.4%, 4.6%, and 2.7% compared to those of the original YOLOv8 model. Additionally, the LAR-YOLOv8 model achieves these improvements while reducing the number of parameters by 40%. The enhanced YOLOv8 [8] method in challenging weather conditions is designed by

employing transfer learning with merged data from different severe weather datasets. The study showed that using customized datasets for training significantly improved the effectiveness of detection compared to the YOLOv8 base values. In addition, the integration of feature-related data with supplementary images constantly enhanced the object detection performance. The deformable convolution (DCN_C2f) module [9], used as the backbone network, and self-calibrating shuffle attention (SC_SA), used for spatial and channel attention processes, are designed for decisions that enable flexible encoding of contextual data, avoiding the loss of feature information that might occur during convolution iterations. It also enhances the ability to represent multiscale, occluded, and tiny object characteristics. The DS-YOLOv8 model exhibits substantial performance enhancements compared to the YOLOv8 series scenarios, surpassing other widely used variants with respect to detection performance. To address the issue of detecting objects at several scales, the SEF module [10] has been developed, which improves upon the SEConv method. Furthermore, the network incorporates an innovative efficient multiscale attention (EMA) mechanism, which is combined with the SPPFE module. The YOLO-SE model has exceptional performance, reaching an average precision of 86.5% at an IoU threshold of 0.5 (AP50) on the SIMD optical remote sensing dataset. To address the problem of recognizing small foreign object debris (FODs), researchers enhanced the improved YOLOv8m [11] model by implementing changes to its structure and adding a specialized, shallow detection component that is specifically designed to accurately identify small items. This approach surpasses YOLOv8m by a difference of 1.02 in AP, resulting in a mean average precision (mAP) of 93.8%. Significantly, improved YOLOv8 outperforms all the anchor-based and anchorless detection methods analyzed, as well as those used in previous research on the FOD-A dataset, in terms of mAP. The image-adaptive YOLO (IA-YOLO) technique [12] is capable of dynamically processing photos under various weather conditions, including both normal and bad weather conditions. The testing results are quite promising, clearly demonstrating the efficacy of the suggested IA-YOLO technique in both hazy and low-light environments.

To aid in search and rescue efforts, Rizk and Bayad [13] presented the efficacy of YOLOv8 deep learning algorithms for person detection in thermal images. The acquired results highlight the flexibility of YOLOv8, which performs well in a variety of circumstances while properly detecting individuals. These findings demonstrate the remarkable 93% precision and 95% AP@50 value of YOLOv8x. The object-detection capabilities of the YOLO algorithm have been effectively utilized in the development of several modules, such as fall, social distance, and tiredness detection [14]. The fall detector serves to enhance the protection of both children and older individuals who are in solitary conditions. The social distance module uses object identification skills to identify individual individuals, followed by the use of bounding boxes to identify potential groups or clusters of individuals. When the distance falls below a certain threshold of 5 pixels, a violation is confirmed. The section can be utilized during pandemics and curfews to effectively control social spaces. The objective of Ghadekar *et al.* [15] was to create a resilient and effective system for identifying suspicious activities in unfavorable weather situations by utilizing the YOLOv7 model the proposed approach aims to identify and categorize potentially dangerous behaviors occurring in challenging environmental conditions, such as rain, fog, and crowded environments. This includes instances of gun usage, physical altercations, and the handling of weapons such as swords. The model has been specifically intended to exhibit strong performance in adverse weather conditions, hence enhancing its practicality in real-world situations. This has been achieved through the utilization of the advanced YOLOv7 architecture. The system's efficacy in consistently identifying and pinpointing dubious activities through comprehensive testing and assessment, even in adverse weather conditions and bustling areas, was investigated.

The study conducted by Kaya *et al.* [16] focused on the automated identification of pedestrian crosswalks within an urban road network, considering the viewpoints of both pedestrians and vehicles. The authors used faster region convolutional neural network (R-CNN) and YOLOv7 model networks to analyze the datasets they obtained. A comparison of the detection performances of the two models revealed that YOLOv7 achieved an accuracy of 98.6%, while Faster R-CNN achieved an accuracy of 98.29%. YOLOv7 outperformed Faster R-CNN in predicting various forms of pedestrian crossings, but the results were relatively similar. Nha *et al.* [17] proposed action and behavior structures as a framework for identifying the necessary components of particular actions in the classroom. We created a database of ten actions-reading, being frustrated, focusing, eating, laughing, using a cell phone, raising a hand, falling asleep, thinking, and writing-based on the suggested framework. Experiments using the YOLO and MobileNetV2 SSD+FPN models yielded positive outcomes. YOLOv7 [18] addresses complex problems by accurately distinguishing between regular pedestrians and electrical system maintenance inspectors. This approach provides a reliable method for quickly detecting and issuing early warnings to pedestrians who may come into contact with power facilities. Researchers developed a unique convolution operation [19] for the activity recognition problem. This operation allows for the heterogeneous exploitation of convolutional filters inside a given layer. To enhance the diversity of the filters, we implement a downsampling operation that modifies the receptive field inside a specific filter group. This adjustment was used to recover the other regular filter groups.

Training deep learning models often requires large datasets specific to the target environment. This limits scalability to new scenarios. CNN often struggle with occlusions and high object density in crowds. Recurrent neural networks (RNNs) might struggle to capture long-range dependencies in complex crowd interactions and offer limited interpretability in their decision-making process. Both CNNs and RNNs can be computationally heavy, especially with large crowds. Long short-term memory networks (LSTMs), or transformers, can capture interactions between individuals but can be computationally expensive. CNN-based methods struggle with identifying individual actions within a crowd. Deep learning models trained on specific datasets can struggle with generalizability to unseen scenarios. This is particularly true for weather variations like rain, snow, or fog, which can significantly alter the visual data. The computational cost of some deep learning methods can hinder their suitability for real-time HAR applications, where fast response is crucial.

YOLOv8 boasts improved object detection, but occlusion handling in dense crowds remains a challenge. While proposed YOLOv8 may adapt well to new scenes, challenges remain when generalizing to highly diverse crowd scenarios (e.g., varying lighting, and crowd composition). This article focuses on YOLOv8 for multi-person activity recognition in crowds under various weather conditions. The main contributions of this research are as follows:

A lightweight, high-precision YOLOv8 detection model is proposed, as shown in Figure 1. This model involves a BiFPN module as the neck layer and ASFF as the head layer to obtain a finer feature representation, addressing the issue of inadequate feature learning extracted from human activity in crowds and the lack of attention given to the impact of human activity features under different weather conditions. The proposed method performs better and is more accurate in terms of human activity detection.

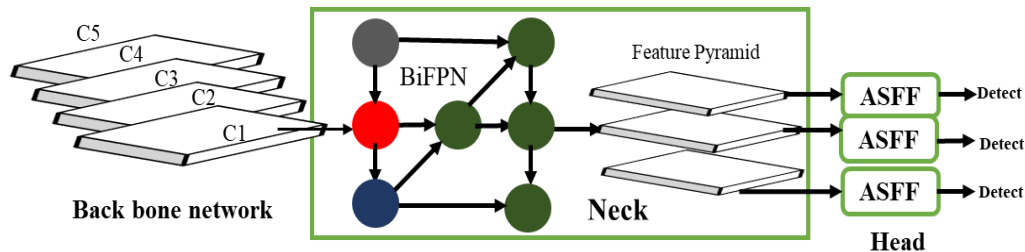


Figure 1. Overall proposed architecture

3. PROPOSED METHOD

YOLOv8, as shown in Figure 2, is an anchor-free, single-stage object detector. We primarily chose this model due to its high accuracy, versatility, ease of deployment and training, and quick detection speed. To further increase its performance and versatility, it incorporates new features and improvements. The three primary advances are a new loss function that can operate on CPUs or GPUs, an anchor-free detector head, and a new backbone network. In contrast to YOLOv5, the C2f structure, as depicted in Figure 3, replaces the C3 frame of the YOLOv5 backbone network. We also adjust distinct channel counts for the different scale models.

The neck structure of YOLOv8 effectively incorporates both the pyramid attention network (PAN) and feature pyramid network (FPN) architectures. The accurate filtering of deep feature data on human activities by the FPN framework provides important, high-level insights at lower levels. A significant amount of crowd feature data that was first received from the YOLOv8 backbone is unintentionally filtered out of the channel data that human activity inputs into the PAN. We dynamically introduced a BiFPN that innovates by incorporating two more lateral connection channels into the existing FPN+PAN structure. The proposed architecture utilizes and preserves the raw properties of human activity to identify the feature map with proficient knowledge, starting from the backbone network. Additionally, the neck layer incorporates a p2 layer. This layer stands out for having a large feature map, few convolution processes, and a second detector head. These enhancements achieve two objectives: they greatly improve microscopic target recognition accuracy and strengthen the model's ability to fuse human position and feature information more successfully.

ASFF is added to YOLOv8's basic feature fusion architecture during the prediction stage to improve model accuracy and ensure that each space can dynamically fuse various feature information levels. The head branch of YOLOv8 uses ASFF to maximize the network's semantic data for higher-level human activity characteristics, such as walking, and spatial data for lower-level activity features, such as standing and running. The BiFPN module generates three feature images and blends them using adaptive weighting.

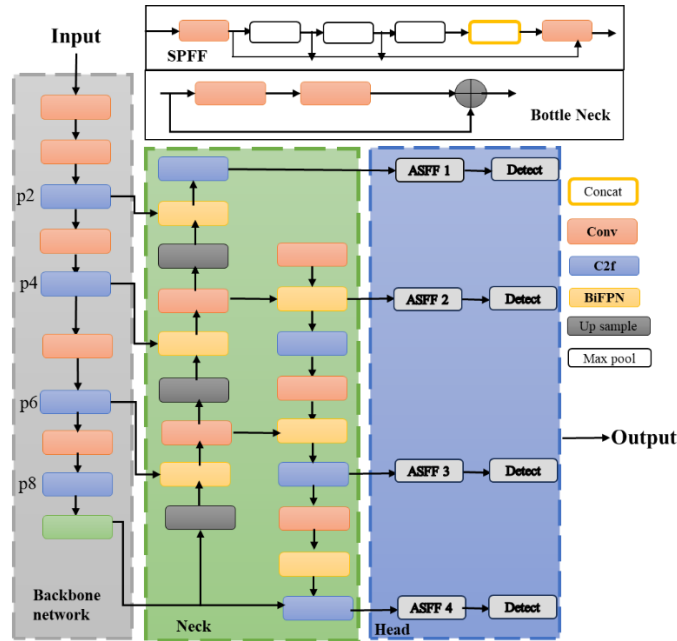


Figure 2. YOLOv8 architecture

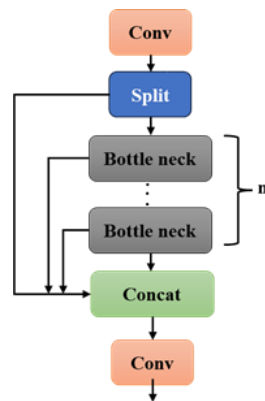


Figure 3. C2f module

4. RESULTS AND DISCUSSION

4.1. Datasets

Publicly available images and videos of crowds under different weather conditions [20] are sourced from platforms such as Google. Analyzing these visual resources allows researchers to assess crowd behavior, response to weather changes, and overall impact on public safety during events. By studying how crowds react to extreme weather conditions [21] such as rain or snow, researchers can better prepare for contingencies and optimize their services accordingly. Additionally, these images and videos serve as useful tools for researchers studying human behavior in diverse environmental settings. Overall, leveraging publicly available imagery from Google offers a wealth of information that can enhance decision-making processes and improve the overall understanding of crowd dynamics under different weather scenarios.

4.2. Experimental setup

An NVIDIA GeForce RTX GPU, Python 3.9, PyTorch 2.0, and CUDA 10.2 are utilized in this experiment. The image input size was 640×640. There were 200 training cycles, and 0.001 was the starting learning rate. The batch size and training iteration count were fixed at 64 and 200, respectively, in every experiment. The augmentations considered are 90°. Rotate clockwise or counterclockwise. Bounding box: Flip: horizontal, vertical.

4.3. Evaluation metrics

Several metrics were taken into consideration in evaluating the effectiveness of human activity detection. Precision measures how many of the activities identified as positive truly are positive. Recall measures how many of the actual positive activities were identified. Average precision (AP) metric considers precision and recall at different IoU thresholds. Mean average precision (mAP) metric averages the AP scores across all human activity categories in the dataset. The equations of evaluation metrics are mentioned in (1)-(4).

$$Precision(P) = \frac{TP}{TP+FP} \quad (1)$$

$$Recall(R) = \frac{TP}{TP+FN} \quad (2)$$

$$AP = \frac{\sum P_{ri}}{\sum r} \quad (3)$$

$$mAP = \frac{AP}{num_classes} \quad (4)$$

4.4. Results visualization

Table 1 lists the precision, recall, and mAP values for the proposed YOLOv8 model under different weather conditions. As demonstrated, incorporating the initial location YOLOv8 with BiFPN and ASFF worked better than other enhanced versions of YOLO and the baseline model at the same time. To help with understanding human activity under different weather conditions as shown in Figure 4. Figure 4(a) represents the human activity in snow condition (37.8%), Figure 4(b) shows the activity during sunlight (87.2%), Figure 4(c) represents dark night detection (40.7%), Figure 4(d) shows activity during fog (15.9%), Figure 4(e) and Figure 4(f) shows the activity in rain (71.9%) and night (LED light) (59.6%) respectively. The confidence score represents the percentage that varies based on the appearance of humans walking with or without occlusion

Table 1. Analysis of YOLOv8 results

Weather conditions	Precision (%)	Recall (%)	mAP (%)
Snow	81.9	32.6	37.8
Sunlight	84.2	82.2	87.2
Fog	66.1	18.8	15.9
Rain	80.5	67.1	71.9
Dark Night	73.1	31.2	40.7
Night (LED light)	61.6	62.9	59.6



Figure 4. HAR results under different weather conditions (a) snow, (b) sunlight, (c) dark night, (d) fog, (e) rain, and (f) dark LED light

Figure 5 represents the loss curves, which are classified as bounding box loss and class loss, i.e., human activity such as walking in different weather conditions, for both training and validation. Box loss quantifies the discrepancy in forecasting the coordinates of bounding boxes. It motivates the proposed model to modify the predicted bounding boxes to align them with the ground truth boxes. The class loss measures the discrepancy in accurately guessing the human activity for every bounding box. It guarantees that the model precisely recognizes the category of the activity. Defocus loss is a specific type of loss component that enhances human activity in different weather situations where images are out of focus or fuzzy. It motivates the model to prioritize enhancing the detection in difficult circumstances.

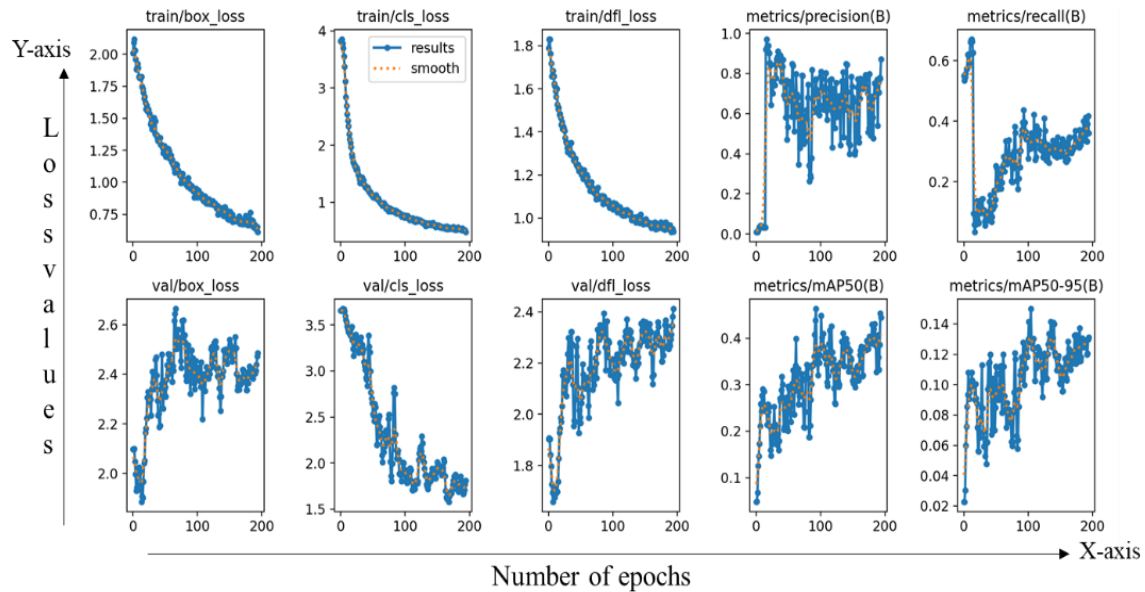


Figure 5. Loss curves for each metric of the proposed method

Calculating the box loss in YOLOv8 involves computing the localization loss and confidence loss, as these are directly related to the bounding box predictions. Actual implementation requires additional considerations depending on specific human actions or optimizations made to the YOLOv8 architecture. To calculate the class loss in YOLOv8, we typically use SoftMax cross-entropy loss, which measures the difference between the predicted human action class probabilities and the ground truth activity class labels. The dynamic feature learning loss (*dfl_loss*) encourages the proposed model to utilize human action features from different scales (weather conditions) and depths of the network, contributing to better human action detection performance, especially for human action in various weather conditions. From the observation, we conclude that for training, the loss steadily decreases when human actions are identified in the video sequence. For validation, there is some variation in loss due to different weather conditions when identifying human action.

4.5. Comparison with other state-of-the-art methods

The classification results using a 2D ViT with LSTM and a 3D ResNet 50 [22] showed promising performance for HMDB51. In the training and testing phases, the accuracy scores were $96.7 \pm 0.35\%$ and $41.0 \pm 0.27\%$, respectively. The architecture combines a ConvLSTM layer with a 3DCNN network component, and it is trained and tested using three databases: LoDVP abnormal activity, MOD20, and UCF50 mini (10 classes) [23]. The combined accuracies of the MOD20, LoDVP abnormal activity, and UCF50mini dataset testing were 78.21%, and 93.471, respectively. The classification phase improved the performance of all CNN models, leading to a reduction in their training time. Using the MSR Pairs dataset, the 3D CNN [24] achieved 100% accuracy when identifying the two datasets based on the types of activities. support vector machine (SVM) using the radial basis function (RBF) kernel algorithm required the least training time (89%), whereas CNN-LSTM required the most training time. When higher precision is needed, it can employ the CNN-LSTM, which achieves an accuracy of 95% using a smartphone dataset [25]. Compared with the above methods, the proposed YOLOv8 achieves an overall accuracy of 96.5% for real-time video and image datasets, as tabulated in Table 2.

The proposed YOLOv8 offers a promising balance between speed and accuracy for HAR. However, weather conditions can affect its performance. Other methods, like LSTMs, might be helpful in specific scenarios, but they come with computational costs. The proposed YOLOv8 is combined with pre-processing techniques, which improve its performance under various weather conditions.

Table 2. Comparison of the proposed method with the literature

Reference	Datasets	Model	Accuracy (%)
[22]	HMDB51	2D ViT + LSTM	41.0
		3D Resnet 50	96.7
[23]	LoDVP dataset	3DCNN + ConvLSTM	93.41
	UCF50 mini dataset		87.78
	MOD20 dataset		78.21
[24]	MSR pairs dataset	3D CNN	100
[25]	Smartphone dataset	LSTM	95
		Logistic regression	90
		SVM	89
	Proposed Method	Real-time videos/images	96.5

5. CONCLUSION

In this work, we propose enhancing YOLOv8-based human activity identification in inclement weather by training using transfer learning with integrated data from many severe weather datasets. We identified significant human activity along the roadside during severe weather using real-time videos and photos. Initially, the first ten items of the COCO datasets were tagged in the YOLOv8 format using the datasets that were gathered from the relevant sources. We used 80% of the individual weather images for training, 5% for validation, and 15% for testing. The meteorological features in the datasets include fog, rain, snow, and darkness. Based on the outcomes of many data augmentations and their identification performances, the best-augmented variant is chosen. The COCO images were used to train custom weights and assess how well they performed in human activity on the testing images. The YOLO method's accuracy table is tabulated as a result of the training procedure, which ended with the achievement of the results on the validation images. Compared to the baseline results of the YOLOv8 algorithm, the proposed method significantly increased the human activity accuracy. The graphed findings demonstrate that the real-time dataset outperformed equal training on the available online datasets. The study concluded that improving the diversity of images from feature-relevant images could improve the prediction of human activities. These findings will benefit future research in a variety of ways. Working with the basic information and labels provided here and adding further datasets could improve the detection even more. A machine with more capacity could achieve even better outcomes by training with more epochs or larger image sizes. Adding additional images from data on typical weather may improve the recognition of human activity beyond extreme weather.

The proposed YOLOv8's focus on speed allows for real-time HAR applications. This is useful for tasks like anomaly detection in surveillance, activity monitoring in healthcare, or human-computer interaction. Compared to previous deep learning methods, the proposed YOLO v8 efficiency could lead to lower computational costs for HAR systems. This could enable deployment on resource-constrained devices. YOLOv8's strong object detection capabilities could be beneficial for large-scale crowd activity recognition. This has applications in crowd management, traffic analysis, and public safety.





REFERENCES

- [1] H. X. Nguyen, D. N. Hoang, H. V. Bui, and T. M. Dang, "Development of a human daily action recognition system for smart-building applications," in *The International Conference on Intelligent Systems & Networks*, 2023, pp. 366–373.
- [2] E. U. Haq, H. Jianjun, K. Li, and H. U. Haq, "Human detection and tracking with deep convolutional neural networks under the constrained of noise and occluded scenes," *Multimedia Tools and Applications*, vol. 79, no. 41–42, pp. 30685–30708, Nov. 2020, doi: 10.1007/s11042-020-09579-x.
- [3] F. M. A. Mazen, R. A. A. Seoud, and Y. O. Shaker, "Deep learning for automatic defect detection in PV modules using electroluminescence images," *IEEE Access*, vol. 11, pp. 57783–57795, 2023, doi: 10.1109/ACCESS.2023.3284043.
- [4] X. Wang, H. Gao, Z. Jia, and Z. Li, "BL-YOLOv8: an improved road defect detection model based on YOLOv8," *Sensors*, vol. 23, no. 20, Oct. 2023, doi: 10.3390/s23208361.
- [5] H. Lou *et al.*, "DC-YOLOv8: small-size object detection algorithm based on camera sensor," *Electronics*, vol. 12, no. 10, May 2023, doi: 10.3390/electronics12102323.
- [6] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, "UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios," *Sensors*, vol. 23, no. 16, Aug. 2023, doi: 10.3390/s23167190.
- [7] H. Yi, B. Liu, B. Zhao, and E. Liu, "Small object detection algorithm based on improved YOLOv8 for remote sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 1734–1747, 2024, doi: 10.1109/JSTARS.2023.3339235.





- [8] D. Kumar and N. Muhammad, "Object detection in adverse weather for autonomous driving through data merging and YOLOv8," *Sensors*, vol. 23, no. 20, Oct. 2023, doi: 10.3390/s23208471.
- [9] L. Shen, B. Lang, and Z. Song, "DS-YOLOv8-based object detection method for remote sensing images," *IEEE Access*, vol. 11, pp. 125122–125137, 2023, doi: 10.1109/ACCESS.2023.3330844.
- [10] T. Wu and Y. Dong, "YOLO-SE: improved YOLOv8 for remote sensing object detection and recognition," *Applied Sciences*, vol. 13, no. 24, Dec. 2023, doi: 10.3390/app132412977.
- [11] J. Farooq, M. Muaz, K. Khan Jadoon, N. Aafaq, and M. K. A. Khan, "An improved YOLOv8 for foreign object debris detection with optimized architecture for small objects," *Multimedia Tools and Applications*, vol. 83, no. 21, pp. 60921–60947, Dec. 2023, doi: 10.1007/s11042-023-17838-w.
- [12] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive YOLO for object detection in adverse weather conditions," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, pp. 1792–1800, Jun. 2022, doi: 10.1609/aaai.v36i2.20072.
- [13] M. Rizk and I. Bayad, "Human detection in thermal images using YOLOv8 for search and rescue missions," in *2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME)*, Oct. 2023, pp. 210–215, doi: 10.1109/ICABME59496.2023.10293139.
- [14] S. Phatangare, S. Kate, D. Khandelwal, A. Khandetod, and A. Kharade, "Real time human activity detection using YOLOv7," in *2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Oct. 2023, pp. 1069–1076, doi: 10.1109/I-SMAC58438.2023.10290168.
- [15] P. Ghadekar, S. Jagtap, B. Sadmak, N. Mane, K. Singh, and B. Chavan, "Suspicious activity detection in adverse weather conditions using YOLOv7," *Grenze International Journal of Engineering & Technology (GIJET)*, 2024.
- [16] Ö. Kaya, M. Y. Çodur, and E. Mustafaraj, "Automatic detection of pedestrian crosswalk with Faster R-CNN and YOLOv7," *Buildings*, vol. 13, no. 4, Apr. 2023, doi: 10.3390/buildings13041070.
- [17] F.-C. Lin, H.-H. Ngo, C.-R. Dow, K.-H. Lam, and H. L. Le, "Student behavior recognition system for the classroom environment based on skeleton pose estimation and person detection," *Sensors*, vol. 21, no. 16, Aug. 2021, doi: 10.3390/s21165314.
- [18] W. Cao, L. Li, S. Gong, and X. Dong, "Research on human behavior feature recognition and intelligent early warning methods in safety supervision scene video based on YOLOv7," *Journal of Physics: Conference Series*, vol. 2496, no. 1, May 2023, doi: 10.1088/1742-6596/2496/1/012019.
- [19] C. Han, L. Zhang, Y. Tang, W. Huang, F. Min, and J. He, "Human activity recognition using wearable sensors by heterogeneous convolutional neural networks," *Expert Systems with Applications*, vol. 198, Jul. 2022, doi: 10.1016/j.eswa.2022.116764.
- [20] Q. Qin, K. Chang, M. Huang, and G. Li, "DENet: detection-driven enhancement network for object detection under adverse weather conditions," *Computer Vision – ACCV 2022*, vol. 13843, 2023, pp. 491–507.
- [21] A. Kerim, U. Celikkan, E. Erdem, and A. Erdem, "Using synthetic data for person tracking under adverse weather conditions," *Image and Vision Computing*, vol. 111, p. 104187, Jul. 2021, doi: 10.1016/j.imavis.2021.104187.
- [22] H. H. Pham, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin, "Video-based human action recognition using deep learning: a review," *arXiv preprint arXiv:2208.03775*, 2022.
- [23] R. Vrskova, P. Kamencay, R. Hudec, and P. Sykora, "A new deep-learning method for human activity recognition," *Sensors*, vol. 23, no. 5, Mar. 2023, doi: 10.3390/s23052816.
- [24] A. Çalışkan, "Detecting human activity types from 3D posture data using deep learning models," *Biomedical Signal Processing and Control*, vol. 81, Mar. 2023, doi: 10.1016/j.bspc.2022.104479.
- [25] B. Moradi, M. Aghapour, and A. Shirbandi, "Compare of machine learning and deep learning approaches for human activity recognition," in *2022 30th International Conference on Electrical Engineering (ICEE)*, May 2022, pp. 592–596, doi: 10.1109/ICEE55646.2022.9827335.

BIOGRAPHIES OF AUTHORS



Shaamili Rajakumar     is currently a research scholar from Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, KTR campus. She has completed MTech embedded systems and BE in ECE. Her research interest is deep learning, computer vision and internet of things. She can be contacted at email: sr6582@srmist.edu.in.



Ruhan Bevi Azad     is an associate professor in the Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, KTR campus. She holds a master's degree in embedded system technologies from Anna University, Chennai and a Ph.D. in security in embedded systems from SRM Institute of Science and Technology. Her research interests include deep learning, image cognition, signal and image processing with machine learning, neural networks, IoT for automation, security in embedded systems, and reconfigurable computing. She has twenty-three years of teaching experience. She can be reached via email at ruhanb@srmist.edu.in.