JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Selfish Herd Optimizer with Deep Learning Driven Object detection and Classification Model for Video Surveillance Systems

[1,*]V. Saikrishnan, [2]Dr. M. Karthikeyan
[1,*]Department of Computer and Information science, Faculty of Science, Annamalai University, Tamil Nadu, INDIA
[2]Department of Computer and Information science, Faculty of Science, Annamalai University, Tamil Nadu, INDIA
*Corresponding Author: V. Saikrishnan

Abstract

Video surveillance performs an essential function in maintaining security and situational awareness through different fields, comprising critical infrastructure, smart cities, and public spaces. The requirement for effectual and robust object detection and classification techniques in video surveillance has considerably developed. This study develops a Selfish Herd Optimizer with deep learning Driven Object detection and Classification Model (SHODL-ODC) for Video Surveillance Systems. The main aim of the SHODL-ODC model is to recognize and classify the existence of objects in the surveillance videos. YOLO-v5 works as the backbone of our model, confirming fast object detection in the surveillance video. For improving the speed and performance of the model, we implement the Nadam optimizer for hyperparameter tuning. Additionally, we present an innovative method for object classification employing Multi-Class Support Vector Machines (MSVM). MSVMs are recognized for their capability to effectively manage multi-class classification tasks, ensuring accurate categorization of objects identified by YOLO-v5. To fine-tune the model's parameters and enhance its overall performance, we developed the Selfish Herd Optimizer (SHO). With comprehensive experimentation and assessment, we establish the model's effectiveness in real-time video surveillance conditions, achieving computational efficiency, classification performance, and higher object detection accuracy. The comparison study represented the improved simulated outcomes of the SHODL-ODC model over the other techniques.

Keywords: Video surveillance; Object detection; Deep learning; Object classification; Parameter tuning; YOLO-v5

1. Introduction

In the past few years, object classification has seen significant growth due to the existing growth in DL-based techniques. Object classification has been explored by various studies in object classification for video as the image to classify objects in video [1]. However, people never contemplate each frame as an independent image in the process of classifying objects into the video, rather than finding the utmost significant effort about a previous frame [2]. Therefore, video classification innovations such as automotive driving, intelligent robotics, and video surveillance are a combination of object tracing and recognition methods [3]. Lately, methods such as region-based convolutional network (R-CNN), single shot detector (SSD), and you only look once (YOLO) have happened that display owing outcomes generally in real-time object classification [4]. However advanced performance hardware is required for utilizing them in the video investigation area.

Object classifications are widely presented in military object recognition, intelligent monitoring, UAV navigation, smart transportation, and unmanned vehicles [5]. However, as an outcome of the several object recognition approaches, the present technique flops to identify the object [6]. Complex background and Changing light raise the difficulty of object classification, precisely for objects in exciting circumstances. The tracing procedure works by identifying an object when it primarily seems in a frame and forecasting its path [7]. This recognition-related technique assesses the object position in each frame separately. It desires an offline teaching technique and could not be proposed to unknown objects. Lately, several recognition and tracing procedures have been presented [8]. However, the stimulating difficulty meets throughout that method means that the arenas need significant study. The difficulty that makes difficult in detection and tracing like dynamic backgrounds, rapid illumination changes, shadow recognition, and moving cameras, so on [9]. Such encounters could not be resolved over the easiest technique due to the complications, inaccuracy, camera jitter, and unconvincing reasons prevailing in the midway stage. To overcome this, computation intelligence (CI) and DL approaches were presented [10].

This article designs a Selfish Herd Optimizer with deep learning Driven Object detection and Classification Model (SHODL-ODC) for Video Surveillance Systems. The major goal of the SHODL-ODC model is to identify and classify the presence of objects present in the surveillance videos. YOLO-v5 serves as the foundation of our model, ensuring rapid object detection in the surveillance feed. For improving the speed of the model and performance, we apply the Nadam optimizer for hyperparameter tuning. Furthermore, we introduce a novel approach for object classification using Multi-Class Support Vector Machines (MSVM). To fine-tune the model's parameters and optimize its overall performance, we introduce the Selfish Herd Optimizer (SHO). The comparison study highlighted the improved simulated outcomes of the SHODL-ODC model over the other techniques.

2. Related works

The author at [11] proposed for teaching a field adoptive scene precise ordinary sensor in unsubstantiated technique. The common sensor has been transferred for different objective fields in a categorized cause field dataset in the absence of human interpreted target model. In depth, it mainly extended the common sensor to double border identification and collected rigid models as untagged objective models related to the classification of self-reliance. Then, devised a sequence semantic transmission technique for arranging the instance level and class level circulations between the objective and basis areas and classifying the solid example habitually. The author at [12] devised a MODT technique. The proposed method uses an optimal Kalman straining technique for tracing the moving object in the video frame. The video slide was changed in accordance with the volume of frames into linguistic procedure over the area increasing technique.

The scientists in [13] proposed an achieving method for motion tracking and object recognition. Now, the authors proposed solid video object tracing and recognition technique. The ambiguous linguistic strainer was done to eliminate the sound present in the front divided structure. The author in [14] proposed a technique called N-YOLO that alternately alters the image stage in YOLO, it separates in permanent magnitude image used in YOLO and joins recognition results of sub-images with interpretation results at different periods relative to tracing method the sum of calculation for object tracing and recognition is significantly decreased. The author in [15], a Contextual Sculpting technique is proposed by a Biased Illumination Field Fuzzy C means a method for identifying the movement of the object exactly. Now, the nonstationary pixel is detached from stationary pixel by employing surrounding Deduction. Then, the Biased Illumination Field FCM method was reached to improve the separation precision by gathering in fluctuating and sound lighting circumstances.

3. The proposed method

A novel SHODL-ODC method has designed for detecting and classifying objects existing in the surveillance videos. The SHODL-ODC model contains MSVM-based classification, YOLO-v5 object detector, SHO based parameter tuning, and Nadam optimizer-based hyperparameter tuning.

3.1. Object Detection Module

The SHODL-ODC method mainly implements YOLO-v5 object detector for efficiently identify the targeted objects. YOLO-v5, an acronym for "You Only Look Once version 5," stands as a pinnacle of real-time object detection in the domain of CV [16]. This is extremely accurate and effective object detection model is re-evaluated the landscape by offering lightning-fast inference speeds with no compromise on accuracy. YOLO-v5 utilizes a streamlined framework to directly forecast class probabilities boxes and bounding, making it remarkably compatible with different applications, from autonomous vehicles to video surveillance. Its accuracy, speed, and versatility design YOLO-v5 the necessary tool in continuously developing the domain of object detection, supporting an extensive of innovative and real-time solutions.

In this phase, Nadam optimizer has been implemented for fine-tuning the hyperparameters of the YOLOv5 framework. The Nadam optimizer, a hybrid of Nesterov Accelerated Gradient (NAG) and Adaptive Moment Estimation (Adam) offers an effective tool while fine-tuning hyperparameters in DL methods [17]. Be excellently combining NAG's momentum-based update with Adam's adaptive learning rates, Nadam gives the benefits of fast convergence and robust optimization. This creates it specifically compatible with hyperparameter tuning, where the balance among exploitation and exploration is important. The optimizer's capability to adaptably modify learning rates for all parameters dependent upon previous gradients confirms that the model's hyperparameters maximum convergence leads to higher accuracy and effective DL techniques.

When implementing to hyperparameter tuning, Nadam overcame the difficulties related to determining the right fusion of hyperparameters to manage a training method. By leveraging its unique integration of NAG's momentum for Adam's adaptability and fast convergence for effective parameter updates, Nadam not only improves the optimization method however, supports to prevention of common pitfalls namely exploding or vanishing gradients. This develops a crucial selection for fine-tuning the hyperparameters of DL technique, finally resulting in enhanced model effectiveness and higher generalization to real-time data.

3.2. Object Classification Module

In object classification method, the MSVM system has employed for assigning class labels. The MSVM classification relies on Vapnik–Chervonenkis (VC) dimension of the statistical learning method. The major aim is to map the pre-processing, non-linear microarray gene expression data to a linear dimensional manifold θ with usage of alteration $\emptyset: R^N \rightarrow \theta$ then, achieving an optimized hyperplane: $\Psi: \psi(x) = (\omega \cdot \phi(x) + b)$ determine the successive augmented convex difficulties (the soft margin issue) [18]:

Determined by

$$y_i(\omega \cdot \phi(x) + b) \geq 1 - \xi_i, for\ all\ 1 \leq i \leq n, \tag{1}$$

Where ω means the coefficient vectors of hyperplane from the manifold, b represents the threshold values of hyperplane, ξ_i indicates the slack difficulty provided to classification error, and β represents the penalty factor to error. The variable β manages the penalty of misclassified and the value was commonly resolved by cross-validation. The higher values of β result in a smaller margin, which decreases classification errors yet, lower values of β produce an extensive margin outcome from different misclassified.

The feature space θ can be a higher dimension; hereafter, the direct evaluation outcomes in "dimension disaster." Next, $\omega = \sum_{i=1}^{n} \delta_i\, y_i \emptyset(x_i)$, all processes of MSVM in the feature space θ was dot product. Subsequently, kernel function $(x_i, x_{i'}) = \emptyset(x_i) \cdot \emptyset(x_{i'})$ that have been efficient at handling dot product, it could be developed to SVM. Thus, the choice of kernel and coefficient is necessary for the computational accuracy as well as the effectiveness of MSVM classifier approaches.

The kernel function can be utilized as a constant predictor:

The linear kernel is given by.

$$G(x_i, x_{i'}) = x_i \cdot x_{i'}. \tag{2}$$

Afterwards, the polynomial kernel is presented in the following equation

$$G(x_i, x_{i'}) = (\eta * (x_i \cdot x_{i'}) + \delta)^d, \tag{3}$$

In cases $> 0, \delta \in R$, and $d \in Z^+$.

Next, the Gaussian kernel is correlated with

$$G(x_i, x_{i'}) = \exp\left(\frac{x_i - x_i^2}{2\sigma^2}\right), \tag{4}$$

If $\sigma > 0$.

The MSVM kernel function is nearly captured into the following: global and local kernel functions. Instance widely various is a major effect on the global kernel values then, sampled close together significantly manage the local kernel values. The linear kernels and polynomial could be optimal instances of global kernels then, the Gaussian and Gaussian radial basis functions (RBF) were local kernels.

3.3. Parameter Tuning Model

The SHO technique has been utilized for tuning the parameters that includes in the MSVM system. SHO has a swarm optimization approach [19]. The population includes two different groups such as Pack predators (P) and Perd of Prey (H). All the members of population can be a survival value calculated by Eq. (5).

$$SV_i = \frac{f_i - f_{best}}{f_{best} - f_{worsi}} \tag{5}$$

In Eq. (5), f_i refers to the fitness value of location of the i^{th} members. f_{best} and f_{worst} denotes the best and worst fitness values. Using the SHO iteration, the herd movement operator was firstly employed for all the members in H that include two different kinds of movement: desertion movement and leader movement herd and following herd. Fig. 1 represents the flowchart of SHO. The leader location of herd viz., best in H, can be updated by the following expression:

$$x_L^{t+1} = \begin{cases} x_L^t + c^t \; if \; SV_L = 1 \\ x_L^t + s^t \; if \; SV_L < 1 \end{cases} \tag{6}$$

In Eq. (6), c^t and s^t are the movement vectors that rely on selfish repulsion and attraction experiments. Within H, the individual members except the leader are separated into groups of herd deserters (H_D) and herd followers (H), and all the individual member moves based on the following equation:

$$x_i^{t+1} = \begin{cases} x_i^t + f_i^t \; if \; i \in H_F^t \\ x_i^t + d_i^t \; if \; i \in H_D^t \end{cases} \tag{7}$$

Now, f_i^t shows the subsequent motion vector and d_i^t refers to the herd desertion vector at t^{th} iteration. Next, the predator movement operator is employed for all the members in P.

$$x_i^{t+1} = x_i^t + 2\rho(x_h^t - x_i^t) \tag{8}$$

Now, x_h^t refers to the location of randomly chosen member in H, ρ shows the random integer within [0,1]. After predation, the iteration of SHO ends and restoration stages are implemented. During the predation stage, the predator kills individual members of the herds with the possibility of being hunted. This member is detached from the population. During the restoration stage, newest herd member is produced from the residual herd member with mating possibility. After these stages, if the stopping condition is met, then iteration is completed.

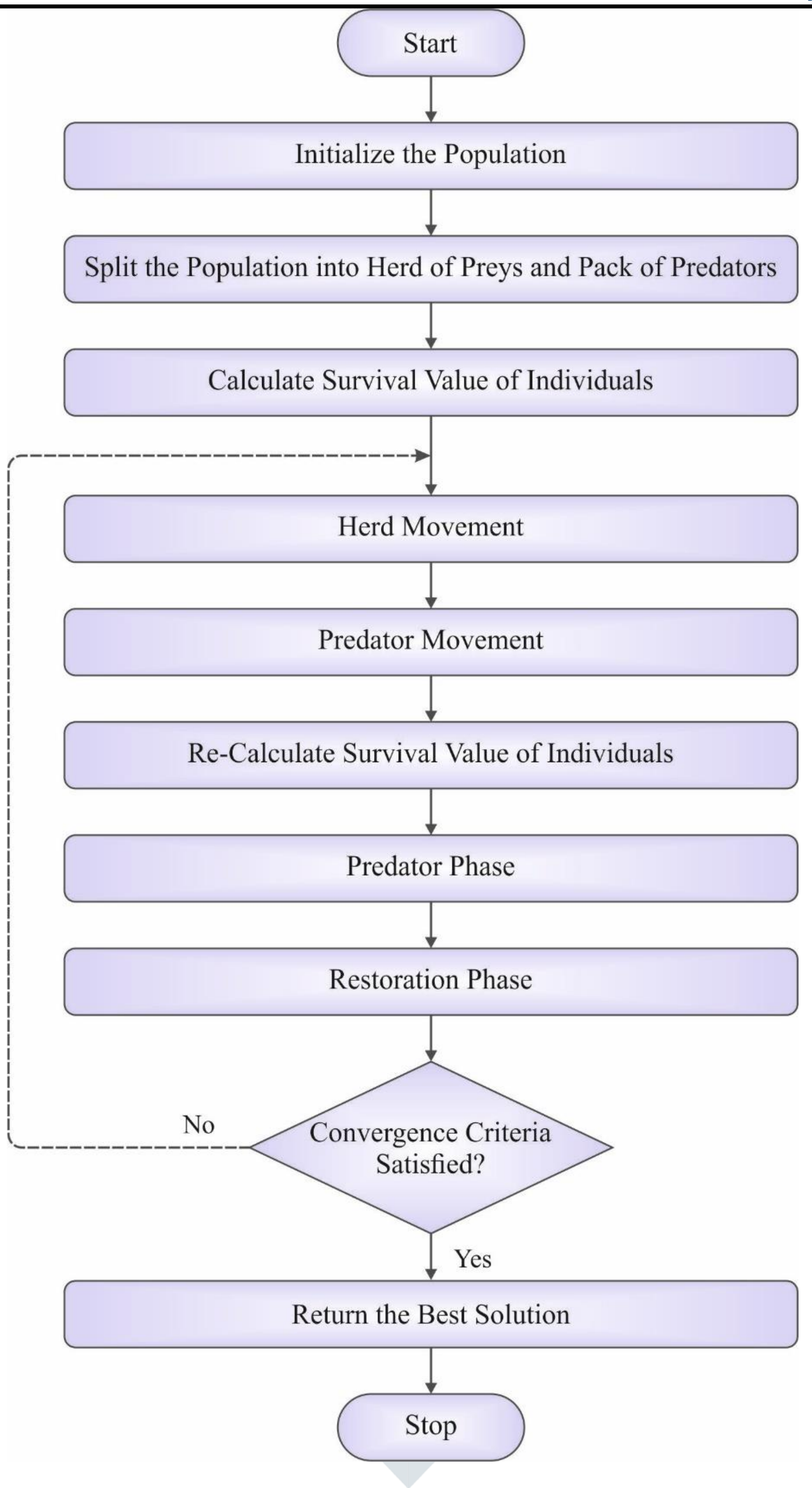

Fig. 1. Flowchart of SHO

4. Performance validation

The performance analysis of the SHODL-ODC method is tested on the UCSDPed2 database [20] that consists of a 2 subsets like pedestrian1 and pedestrian2 datasets. Table 1 represents the dataset description. Fig. 2 exhibits the sample test image with corresponding ground truth images.

Table 1 Description of dataset

Dataset	Testbed	Frames No.	Time (sec)
UCSDped2	Pedestrian1 Dataset	360	12
	Pedestrian2 Dataset		

(a) (b)

Fig. 2. a) Original Image b) Ground Truth Image

Table 2 and Fig. 3 emphasize the average detection accuracy of the SHODL-ODC method on a 2 databases. The figure shows that the SHODL-ODC technique has attained higher performance over the other systems on 2 sub-databases. For a case, with a surveillance ped-1 database, the SHODL-ODC technique is provided better average accuracy of 98.21% while the CIHSART-ODT, DLA-DT, Region-CNN, and FR-CNN methods acquired lesser average accuracy of 98%, 97%, 97%, and 85%. Besides, with the surveillance ped-2 dataset, the SHODL-ODC model can be given improved average accuracy of 92.92% but the CIHSART-ODT, DLA-DT, Region-CNN, and FR-CNN algorithms obtained a reduced average accuracy of 91%, 90%, 87%, and 82%.

Table 2 Average $accu_y$ the outcome of Analysis of SHODL-ODC method on two datasets

Methods	**SHODL-ODC**	**CIHSART-ODT**	**DLA-DT**	**Region-CNN**	**FR-CNN**
Surveillance Ped-1	98.21	98.00	97.00	97.00	85.00
Surveillance Ped-2	92.92	91.00	90.00	87.00	82.00

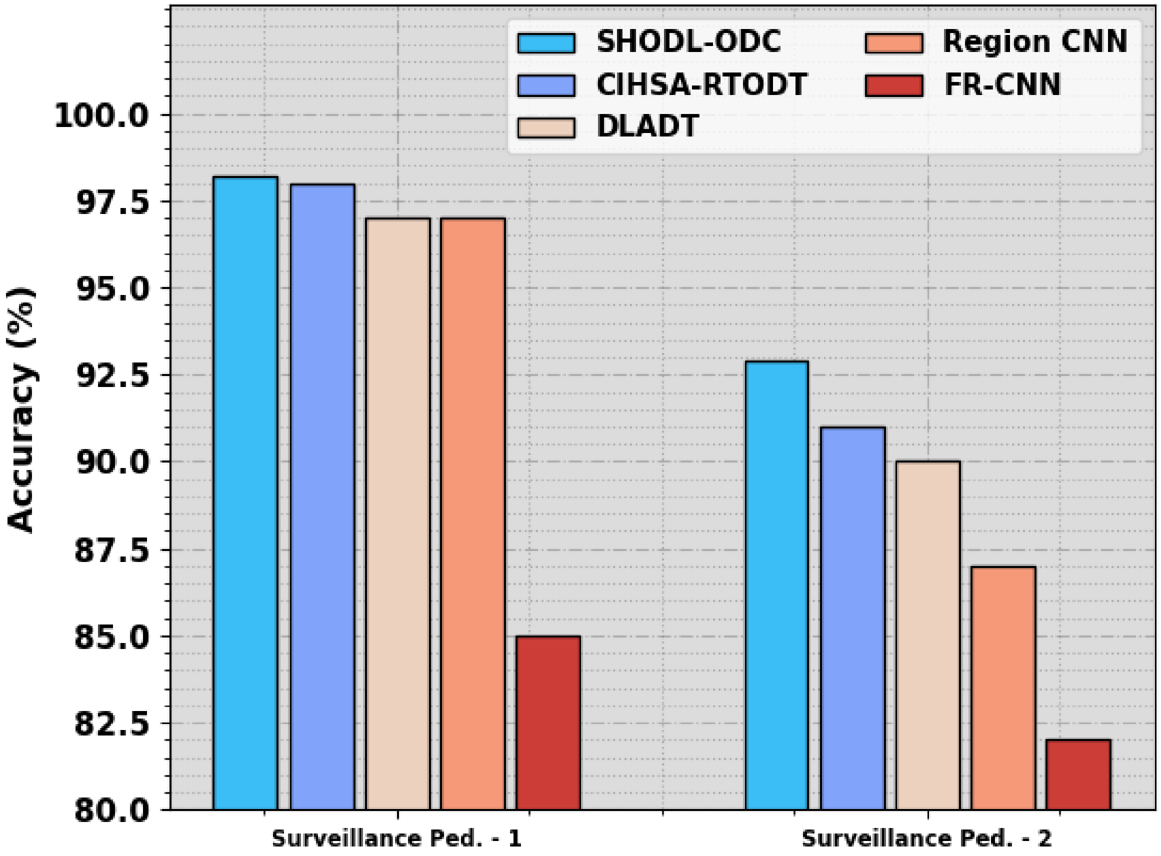

Fig. 3. Average accuracy of SHODL-ODC model on two sub-datasets

Table 3 and Fig. 4 analysis of the AUC of the SHODL-ODC technique on a 2 sub-datasets [21-24]. The figure shows that the SHODL-ODC model achieved higher performance over the other techniques on 2 sub datasets. Based on, with surveillance ped-1 database, the SHODL-ODC system gets better AUC of 97.85% while the MP-PCA, SF, SFMP-PCA, M-DT, A-MDN, AD-VAE, and CIHSART-ODT methodologies acquired lesser AUC of 61.01%, 66.74%, 67.25%, 82.05%, 91.71%, 95.39%, and 97.12%. Besides, with surveillance ped-2 database, the SHODL-ODC approach is provided maximum AUC of 94.74% whereas the MP-PCA, SF, SFMP-PCA, M-DT, A-MDN, AD-VAE, and CIHSART-ODT models are obtained the decreased value AUC of 69.92%, 55.96%, 61.33%, 82.99%, 91.25%, 92.47%, and 93.92%.

Table 3 AUC analysis of SHODL-ODC model on two sub datasets

Models	**Surveillance Ped.1**	**Surveillance Ped.2**
MP-PCA	61.01	69.92
SF	66.74	55.96
SFMP-PCA	67.25	61.33
M-DT	82.05	82.99
A-MDN	91.71	91.25
AD-VAE	95.39	92.47
CIHSART-ODT	97.12	93.92
SHODL-ODC	97.85	94.74

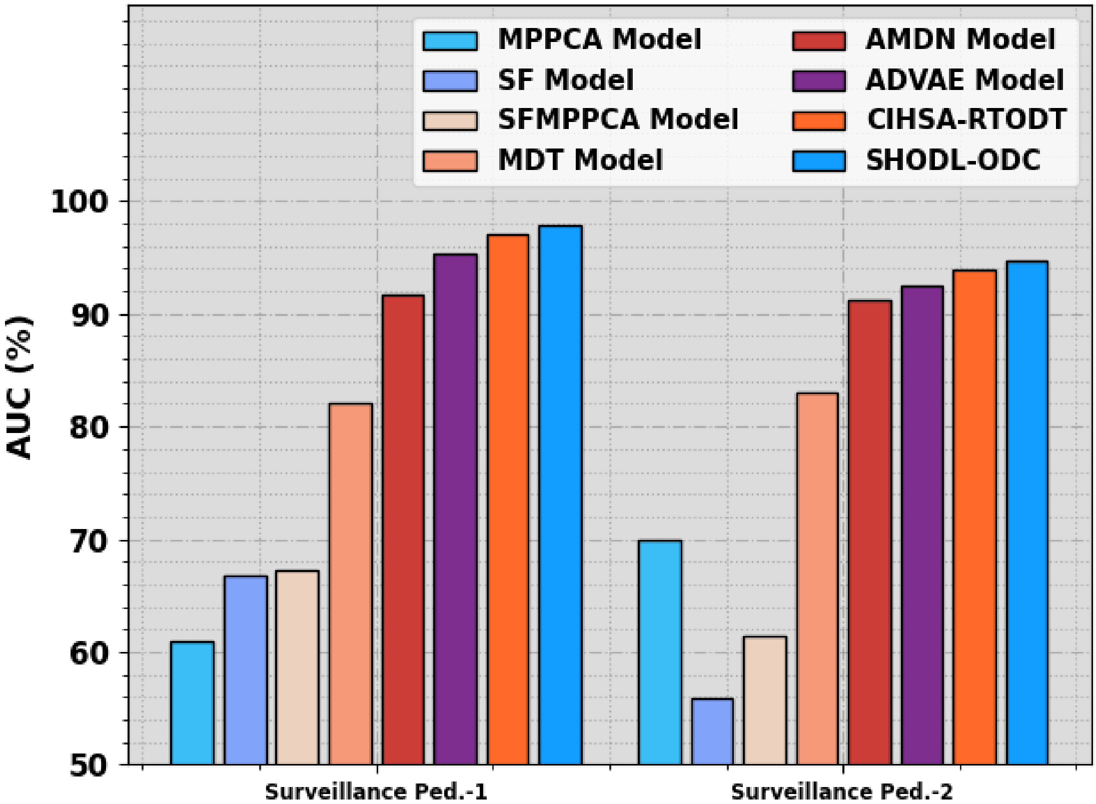

Fig. 4. AUC analysis of SHODL-ODC model on two sub datasets

Table 4 and Fig. 5 illustrate a running time (RT) analysis of the SHODL-ODC technique over the other models. The simulated values exhibited that SHODL-ODC system gives lesser RT over the other RT. For instance, with surveillance ped-1 dataset, the SHODL-ODC methodologies are given decreased RT of accuracy of 1.84s while the M-DT, SCLF, A-MDN, AD-VAE, and CIHSART-ODT systems obtained minimum RT of 20.61s, 20.11s, 11.73s, 3.94s, and 2.67s correspondingly. Concurrently, with surveillance ped-2 dataset, the SHODL-ODC system offers decreased RT of accuracy of 1.90s whereas the M-DT, SCLF, A-MDN, AD-VAE, and CIHSART-ODT models have acquired lesser values RT of 22.94s, 18.48s, 13.02s, 6.16s, and 3.98s individually.

Table 4 RT analysis of SHODL-ODC model with other approaches on two sub-datasets

Methods	Pedestrian1	Pedestrian2
M-DT	20.61	22.94
SCLF	20.11	18.48
A-MDN	11.73	13.02
AD-VAE	3.94	6.16
CIHSART-ODT	2.67	3.98
SHODL-ODC	1.84	1.90

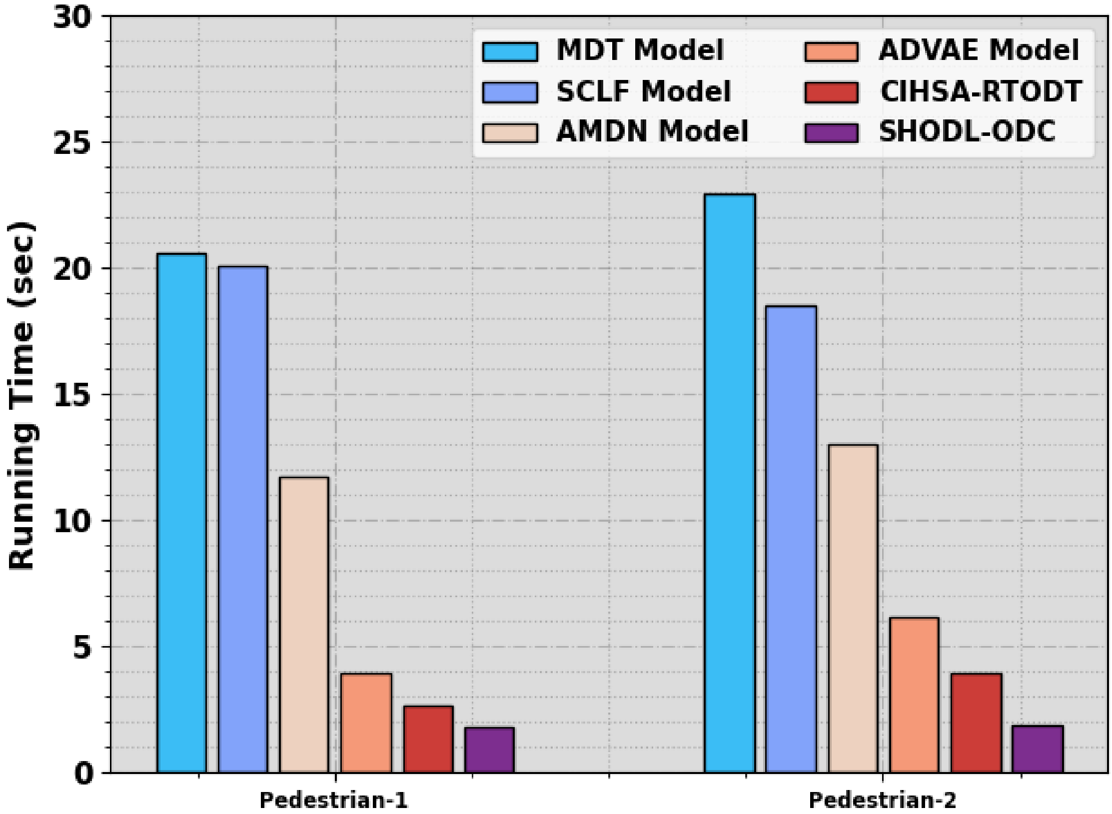

Fig. 5. RT analysis of SHODL-ODC model on two sub datasets

5. Conclusion

In this study, a new SHODL-ODC methodology is developed for the detection and classification of objects present in the surveillance videos. YOLO-v5 serves as the foundation of our model, ensuring rapid object detection in the surveillance feed. To improve the speed of the model and performance, we apply the Nadam optimizer for hyperparameter tuning. Furthermore, we introduce a novel approach for object classification using MSVM. To fine-tune the model's parameters and optimize its overall performance, we introduce the SHO. The comparison analysis noted the improved simulated results of the SHODL-ODC system over the other methodologies.

References

[1] Jha, S., Seo, C., Yang, E. and Joshi, G.P., 2021. Real time object detection and trackingsystem for video surveillance system. Multimedia Tools and Applications, 80(3), pp.3981-3996.

[2] Xu, J., 2021. A deep learning approach to building an intelligent video surveillance system. Multimedia Tools and Applications, 80(4), pp.5495-5515.

[3] Naik, U.P., Rajesh, V. and Kumar, R., 2021, September. Implementation of YOLOv4 Algorithm for Multiple Object Detection in Image and Video Dataset using Deep Learning and Artificial Intelligence for Urban Traffic Video Surveillance Application. In 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-6). IEEE.

[4] Wang, Y., Mao, K., Chen, T., Yin, Y., He, S. and Chen, G., 2021. Accelerating real-time object detection in high-resolution video surveillance. Concurrency and Computation: Practice and Experience, p.e6307.

[5] Gautam, A. and Singh, S., 2021. Deep Learning Based Object Detection Combined with Internet of Things for Remote Surveillance. Wireless Personal Communications, 118(4), pp.2121-2140.

[6] Xu, J., 2021. A deep learning approach to building an intelligent video surveillance system. Multimedia Tools and Applications, 80(4), pp.5495-5515.

[7] Gautam, A. and Singh, S., 2021. Deep Learning Based Object Detection Combined with Internet of Things for Remote Surveillance. Wireless Personal Communications, 118(4), pp.2121-2140.

[8] Sung, C.S. and Park, J.Y., 2021. Design of an intelligent video surveillance system for crime prevention: applying deep learning technology. Multimedia Tools and Applications, 80(26), pp.34297-34309.

[9] Salido, J., Lomas, V., Ruiz-Santaquiteria, J. and Deniz, O., 2021. Automatic handgun detection with deep learning in video surveillance images. Applied Sciences, 11(13), p.6085.

[10] Jiao, L., Zhang, R., Liu, F., Yang, S., Hou, B., Li, L. and Tang, X., 2021. New generation deep learning for video object detection: A survey. IEEE Transactions on Neural Networks and Learning Systems.

[11] Mou, Q., Wei, L., Wang, C., Luo, D., He, S., Zhang, J., Xu, H., Luo, C. and Gao, C., 2021. Unsupervised domain-adaptive scene-specific pedestrian detection for static video surveillance. Pattern Recognition, 118, p.108038.

[12] Elhoseny, M., 2020. Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems. Circuits, Systems, and Signal Processing, 39(2), pp.611-630.

[13] Mahalingam, T. and Subramoniam, M., 2020. A robust single and multiple moving object detection, tracking and classification. Applied Computing and Informatics.

[14] Jha, S., Seo, C., Yang, E. and Joshi, G.P., 2021. Real time object detection and trackingsystem for video surveillance system. Multimedia Tools and Applications, 80(3), pp.3981-3996.

[15] Kalli, S., Suresh, T., Prasanth, A., Muthumanickam, T. and Mohanram, K., 2021. An effective motion object detection using adaptive background modeling mechanism in video surveillance system. Journal of Intelligent & Fuzzy Systems, (Preprint), pp.1-13.

[16] Kuznetsova, A., Maleva, T. and Soloviev, V., 2020, December. Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images. In International Symposium on Neural Networks (pp. 233-243). Springer, Cham.

[17] Xiao, B., Liu, Y. and Xiao, B., 2019. Accurate state-of-charge estimation approach for lithium-ion batteries by gated recurrent unit with ensemble optimizer. Ieee Access, 7, pp.54192-54202.

[18] Thirumala, K., Pal, S., Jain, T. and Umarikar, A.C., 2019. A classification method for multiple power quality disturbances using EWT based adaptive filtering and multiclass SVM. Neurocomputing, 334, pp.265-274.

[19] Kaya, E., Baştemur Kaya, C., Bendeş, E., Atasever, S., Öztürk, B. and Yazlık, B., 2023. Training of Feed-Forward Neural Networks by Using Optimization Algorithms Based on Swarm-Intelligent for Maximum Power Point Tracking. *Biomimetics*, *8*(5), p.402.

[20] http://www.svcl.ucsd.edu/projects/anomaly/dataset.html

[21] Pustokhina, I.V., Pustokhin, D.A., Vaiyapuri, T., Gupta, D., Kumar, S. and Shankar, K., 2021. An automated deep learning based anomaly detection in pedestrian walkways for vulnerable road users safety. Safety Science, 142, p.105356.

[22] Xu, M., Yu, X., Chen, D., Wu, C. and Jiang, Y., 2019. An efficient anomaly detection system for crowded scenes using variational autoencoders. Applied Sciences, 9(16), p.3337.

[23] Murugan, B.S., Elhoseny, M., Shankar, K. and Uthayakumar, J., 2019. Region-based scalable smart system for anomaly detection in pedestrian walkways. Computers & Electrical Engineering, 75, pp.146-160.

[24] Alotaibi, M.F.; Omri, M.; Abdel-Khalek, S.; Khalil, E.; Mansour, R.F. Computational Intelligence-Based Harmony Search Algorithm for Real-Time Object Detection and Tracking in Video Surveillance Systems. Mathematics 2022, 10, 733. https://doi.org/ 10.3390/math10050733.