Automated Object Detection and Classification using Metaheuristics with Deep Learning on Surveillance Videos

V. Saikrishnan[1], Dr. M. Karthikeyan[2]
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

Abstract—**Object detection and classification from surveillance videos is a procedure of exploiting computer vision (CV) and machine learning (ML) techniques for categorizing and identifying objects from realtime video streams. This assists in understanding and analyzing the scene, tracking object movement, and identifying potential security threats. This technology has been extremely utilized in security and surveillance schemes, autonomous vehicles, and retail analytics. This study develops an Automated Object Detection and Classification using Metaheuristics with Deep Learning on Surveillance Videos (AODC-MDLSV) technique. The presented AODC-MDLSV technique proficiently detects and classifies the objects into multiple classes. To do so, the AODC-MDLSV technique initially performs object detection using YOLO-v5 model. In the next stage, the AODC-MDLSV technique employs random vector functional link network (RVFL) method for object classification purposes. Finally, artificial bee colony optimization (ABC) methodology is used as a parameter optimization approach to improve the detection efficiency. The simulation values of the AODC-MDLSV approach are tested on benchmark video and the results showcased the better performance of the AODC-MDLSV method with maximum accuracy of 99.93%.**

Keywords— Video surveillance; Object detection; Object classification; Deep learning; Bee colony optimization

I. INTRODUCTION

Video surveillance refers to a process of evaluating video sequences and it remains to be an active area in computer vision (CV) [1]. It grants immense volume of data storage and display. There exist 3 kinds of Video surveillance activities. Video surveillance activities can be fully-autonomous, manual, or semi-autonomous [2, 3]. Manual video surveillance means as the name itself defines that videos are analyzed by the human. These systems were broadly used whereas semi-autonomous video surveillance is certain form of video processing but with human interference [4]. Object classifications in video series are a continuous developing area which has a great number of applications in various domains like robot navigation, biomedical imaging, video surveillance, remote sensing, vehicle navigation, biometry, and visual inspection [5, 6]. Owing to the rapid development and accessibility of high-quality cameras in video capture technology, video is now been a cheap source of data. This leads to a great interest in the object classification and analysis of video sequences [7]. The different steps indulged in object classification were feature extraction, pre-processing, object detection, classification based on features extracted and conversion of videos into frames. Object classification in videos refers to a complicated process which necessitates highly accurate and robust approaches [8].

Deep learning (DL) based object detection methods are a type of CV technique that leverages deep neural network (DNN) for classifying and detecting objects in videos or images. Such methods were trained on large amounts of data from labelled images for learning the patterns and features that describe various objects [9]. The common DL related object detection methods were depending on convolutional neural network (CNN), which can be devised for processing image datasets. Such techniques commonly use an integration of classification, feature extraction, and object proposal generation to identify objects in an image [10].

Mou et al. [11] presented for training domain-adoptive scene-specific pedestrian detectors in unsupervised approach. The generic sensor was relocated for different targeted fields in one labeled source domain dataset without human-annotated targeted instances. In detail, it can be mainly expanded generic detectors to dual-boundary classification and collected hard sample as unlabelled targeted sample depends on the recognition confidence. Elhoseny [12] devised a MODT approach. The videos are converted as per the amount of frames into morphological operation through region growing approach. The authors in [13] devised an effective technique for object identification and movement tracking. In this study, the author offered strong video object identification and tracking approach. The fuzzy morphological filter was executed for eradicating the noises presented in foreground segmented frames. Jha et al. [14] offered a N-YOLO technique that rather than resizing image step in YOLO, it splits into fixed size imageries leveraged in YOLO and combines identification outcomes of sub-images with inference outcomes at distinct times with relation-oriented tracking method the sum of calculation for object identiifcation and tracking is diminished. In [15], a Background Modelling approach was offered by a Biased Illumination Field Fuzzy C-means method to identify moving objects precisely. Now, the non-stationary pixel was detached from stationary pixel with the use of Background Subtraction. Then, the Biased Illumination Field Fuzzy C-means technique was gained for enhancing segmented precision by clustering in changing and noise illumination situations.

This study develops an Automated Object Detection and Classification using Metaheuristics with Deep Learning on Surveillance Videos (AODC-MDLSV) technique. The presented AODC-MDLSV technique proficiently detects and classifies the objects into multiple classes. To do so, the AODC-MDLSV technique initially performs object detection utilizing YOLO-v5 method. In the next stage, the AODC-MDLSV technique employs random vector functional link network (RVFL) method for object classification purposes. Finally, artificial bee colony optimization (ABC) algorithm was used as a parameter optimization approach to improve the detection efficiency. The result analysis of the AODC-MDLSV approach is tested on benchmark video.

II. THE PROPOSED MODEL

In this study, we have developed a new AODC-MDLSV approach for object detection and classification on surveillance videos. The presented AODC-MDLSV technique proficiently detects and classifies the objects into multiple classes. To do so, the AODC-MDLSV technique performs YOLO-v5 object detection, RVFL classification, and ABC based parameter optimization.

A. *Object Detection using YOLO-v5*

The AODC-MDLSV technique initially performs object detection using YOLO-v5 model. Object detection is an essential issue in the domain of computer vision (CV) [16]. It permits the computer to discover and locate designation of interests from an image automatic such as flaws in the fabric. DL-based object detection systems can be attained great success recently. However, the previous techniques can work to meet the realtime necessity of fabric defect detection method as it can be higher computational overhead. In order to balance precision and speed, a lightweight object detection network termed YOLOv5 was employed under this case. Fig. 1 illustrates the structure of YOLO-V5. The standard YOLOv5 is improved depends on the feature of fabric defects such that it is applied for fabric defect detection approach. The framework of standard YOLOv5 mostly contains Backbone, PANet, and Output. The backbone can employed to execute feature engineering from input image. The PANet take reached visual features robust for scale changing due to exploited pyramid design. The locations are output, and the ROI is classifier simultaneously.

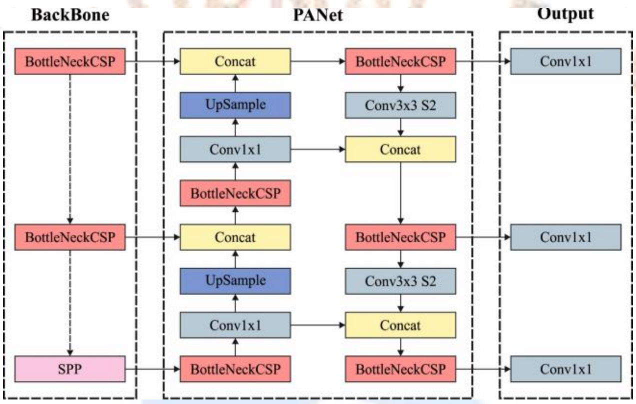

Fig. 1. YOLO-V5 Architecture [16]

B. *Object Classification using RVFL Model*

The AODC-MDLSV technique employed the RVFL model for object classification purposes. A fundamental structure of RVFL was generally related to ANN which contains hidden, output, and input layers [17]. But, huge difference amongst RVFL and ANN was that RVFL directly links output and input layers. This connection assists RVFL with suitable processes for preventing the overfitting issues which occur in ANN. The trained RVFL starts with utilization of input dataset which contains target y_i and instance x_i, to input state. Afterward, the resultant of hidden node was estimated by subsequent equation:

$$O_j(\alpha_j x_i + \beta_j) = \frac{1}{1 + e^{-(\alpha_j x_i + \beta_j)}}, \beta_j \in [0, S], \alpha_j \in [-S, S] \quad (1)$$

In the above formula, α_j signifies the value of weighted that links hidden and input nodes. S and β_j signifies scale bias and factor. Afterwards, the final result was computed by Eq. (11):

$$Z = Fw, w \in R^{n+P}, F = [F_1, F_2] \quad (2)$$

$$F_1 = \begin{bmatrix} x_{11} & x_{1n} \\ \vdots & \vdots \\ x_{N1} & x_{Nn} \end{bmatrix} F_2 = \begin{bmatrix} G_1(\alpha_1 x_1 + \beta_1) & \cdots & G_P(\alpha_P x_1 + \beta_P) \\ \vdots & \ddots & \vdots \\ G_1(\alpha_1 x_N + \beta_1) & \cdots & G_P(\alpha_P x_N + \beta_P) \end{bmatrix} \quad (3)$$

Afterward, w can be upgrading as:

$$w = F^{\dagger} Z \quad (4)$$

In Eq. (4), † designates the Moore-Penrose pseudo-inverse.

C. *Parameter Tuning using ABC Algorithm*

In this work, the ABC algorithm is used as a parameter optimization approach to improve detection efficiency [18]. Dervish Karaboga (2005) coined an ABC algorithm for resolving optimization issues which have gained popularity amongst researcher workers because of its higher efficiency, robustness, and attractive mechanisms. In this work, employed, onlooker, and scout bees are the 3 classifications of ABC. In the 1st half, the employed bee (EB) works whereas onlooker bee plays their role in second half. The amount of food sources is equivalent to the amount of EBs since there exists single EB for all the food sources. As soon as the food sources are applied, and onlooker bee was consumed, then scout bee was made. Briefly, there exist 3 phases in all the cycles of ABC. The initial phase includes directing the EBs for the nectar amount measurement. Next, the EBs transfer the position and nectar amount of foods with onlooker bee where the food source is designated. Then, the purpose of scout bees and sending them to potential food sources are included.

The location of the food source equivalent to the possible solution for optimized problems, whereas fitness of solution was specified through nectar amount. The amount of solutions determine onlooker bee and EBs amount. Initially, ABC generates a population randomly that undergoes recurrent cycle of search from scout, employed, and onlooker bees. Afterward, EBs, complete the search method, they share nectar data of food sources and the corresponding position to the onlooker bee. Next assessment of nectar data can be performed by the onlooker bee, and, adjustments of the location can be made. The collection of food sources by the onlooker bee depends on P_i probability values that can be evaluated as follows:

$$P_i = \frac{FiT_i}{\sum_{n=1}^{N} Fit_n} \quad (5)$$

Let FiT_i signifies fitness value of i^{th} solution and N denotes amount of food resources. To generate the candidate location, subsequent formula was applied

$$V_{ij} = X_{ij} + (\beta_{ij} \times (X_{ij} - X_{kj}) \quad (6)$$

Where $k \in \{1,2,.,EB\}$ and $j \in \{1,2,.,D\}$ are arbitrarily selected index. Now, EB and D represent the quantity of EB s and optimization variables correspondingly.

From the expression, k is randomly chosen, however, it must be inconsistent with i. The $(\beta_{ij}$ denotes a randomly number within $[-1,1]$ interval which control the generation of neighboring food source location around X_{ij}. Afterward generating a candidate location of source, V_{ij} is produced and compared to X_{ij}. According to the assessment, when the nectar amount in novel food source is higher then it is succeeded in the memory of scout bees, or else, keep the earlier one. In another word, the selection process applied was a greedy selection method between previous and current food sources.

The fitness selective is a vital feature from the ABC technique. The solution encoder was employed to assess the aptitude of candidate results. Here, the accurateness value is an important criteria exploited to design a fitness function.

$$Fitness = \max(P) \quad (7)$$

$$P = \frac{TP}{TP + FP} \quad (8)$$

But, TP denotes the true positive and FP represents false positive value.

III. Results and Discussion

The proposed model is simulated using Python 3.6.5 tool on PC i5-8600k, GeForce 1050Ti 4GB, 16GB RAM, 250GB SSD, and 1TB HDD. The performance validation of the AODC-MDLSV method is tested by utilizing the UCSDPed2 dataset [19], including two subsets they pedestrian-1 and pedestrian-2 databases. Table 1 shows the dataset description. Fig. 2 visualizes the sample outcomes of the AODC-MDLSV method. The figure indicated that the AODC-MDLSV method has detected all the objects that present in the video frame.

TABLE I

DETAILS OF DATABASE

Dataset	Testbed	Frames No.	Time (sec)
UCSDped2	Pedestrian-1 Database	360	12
	Pedestrian-2 Database		

Table 2 and Fig. 3 highlight the average detection accuracy of the AODC-MDLSV model on two sub datasets. The figure exhibited the AODC-MDLSV approach has reached maximum performance over the other approaches on two sub datasets. For example, with surveillance ped-1 database, the AODC-MDLSV methodology has offered higher average accuracy of 98.84% where the CIHSA-RTODT, DLADT, Region CNN, and FR-CNN models have obtained lower average accuracy of 98%, 97%, 97%, and 85%. Also, with surveillance ped-2 dataset, the AODC-MDLSV method has offered higher average accuracy of 94.13% where the CIHSA-RTODT, DLADT, Region CNN, and FR-CNN approaches have gained lesser average accuracy of 91%, 90%, 87%, and 82%.

Fig. 2. a) Original Image b) Tracking Image

TABLE II

AVERAGE ACCURACY OUTCOME OF AODC-MDLSV APPROACH ON TWO SUB DATASETS

Methods	AODC-MDLSV	CIHSA-RTODT	DLADT	Region CNN	FR-CNN
Surveillance Ped. - 1	98.84	98.00	97.00	97.00	85.00
Surveillance Ped. - 2	94.13	91.00	90.00	87.00	82.00

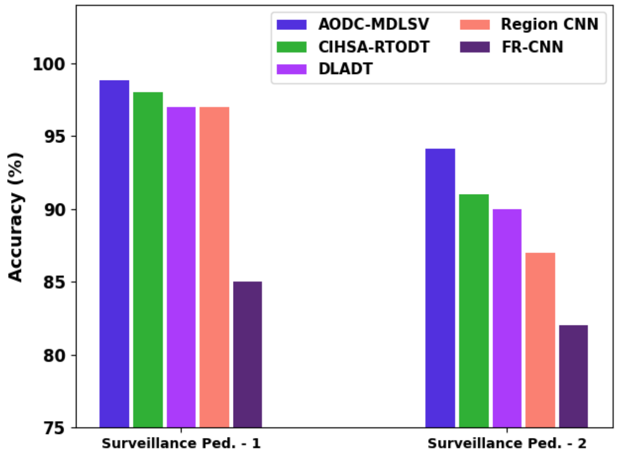

Fig. 3. Average accuracy of AODC-MDLSV system on two sub datasets

Table 3 and Fig. 4 examine the AUC of the AODC-MDLSV system on two sub datasets. The figure represented that the AODC-MDLSV system has reached maximal performance over the other approaches on two sub datasets. For instance, with surveillance ped-1 dataset, the AODC-MDLSV approach has presented higher AUC of 98.98% where the MPPCA, SF, SFMPPCA, MDT, AMDN, ADVAE, and CIHSA-RTODT models have obtained lower AUC of 61.01%, 66.74%, 67.25%, 82.05%, 91.71%, 95.39%, and 97.12%. Likewise, with surveillance ped-2 dataset, the AODC-MDLSV model has offered higher AUC of 95.97% where the MPPCA, SF, SFMPPCA, MDT, AMDN, ADVAE, and CIHSA-RTODT methods have gained minimum AUC of 69.92%, 55.96%, 61.33%, 82.99%, 91.25%, 92.47%, and 93.92%.

TABLE III

AUC ANALYSIS OF AODC-MDLSV ALGORITHM WITH OTHER SYSTEMS ON TWO SUB DATASETS

Models	Surveillance Ped.-1	Surveillance Ped.-2
MPPCA	61.01	69.92
SF	66.74	55.96
SFMPPCA	67.25	61.33
MDT	82.05	82.99
AMDN	91.71	91.25
ADVAE	95.39	92.47
CIHSA-RTODT	97.12	93.92
AODC-MDLSV	98.98	95.97

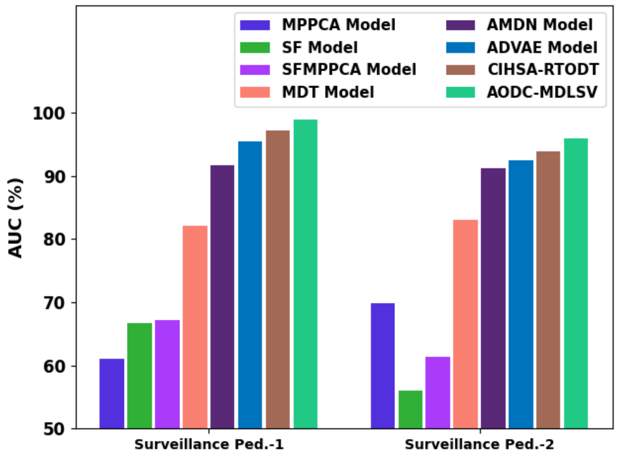

Fig. 4. AUC analysis of AODC-MDLSV approach on two sub datasets

TABLE IV

RT OUTCOME OF AODC-MDLSV APPROACH WITH OTHER SYSTEMS ON TWO SUB DATASETS

Models	Pedestrian-1	Pedestrian-2
MDT Model	20.61	22.94
SCLF Model	20.11	18.48
AMDN Model	11.73	13.02
ADVAE Model	3.94	6.16
CIHSA-RTODT	2.67	3.98
AODC-MDLSV	2.01	2.32

Table 4 and Fig. 5 provide a running time (RT) examination of the AODC-MDLSV method over the other techniques. The outcomes shows the AODC-MDLSV model has provided lower RT over the other techniques. For example, with surveillance ped-1 dataset, the AODC-MDLSV model has provided minimal RT of accuracy of 2.01s where the MDT, SCLF, AMDN, ADVAE, and CIHSA-RTODT methods have established lower RT of 20.61s, 20.11s, 11.73s, 3.94s, and 2.67s respectively. At the same time, with surveillance ped-2 dataset, the AODC-MDLSV approach has provided minimal RT of accuracy of 2.32s where the MDT, SCLF, AMDN, ADVAE, and CIHSA-RTODT methods have lower RT of 22.94s, 18.48s, 13.02s, 6.16s, and 3.98s correspondingly.

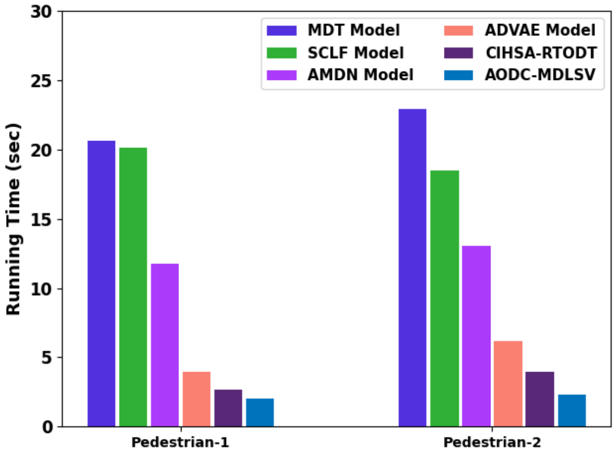

Fig. 5. RT analysis of AODC-MDLSV algorithm on two sub datasets

Table 5 and Fig. 6 highlight the comparative result examination of the AODC-MDLSV model with existing models on surveillance Ped.-1 dataset [6, 20-22]. The figure designated that the AODC-MDLSV method has accomplished maximum ROC values. The outcomes exhibited the SF method has showcased poor outcomes with lower ROC values. Followed by, the ADVAE and CIHSA-RTODT models have demonstrated slightly enhanced ROC values. In line with, the AMDN model has accomplished reasonable ROC values. However, the AODC-MDLSV model has reached maximum performance over the other methods.

TABLE V

COMPARATIVE OUTCOME OF AODC-MDLSV APPROACH WITH EXISTING SYSTEMS ON SURVEILLANCE PED.-1

ROC	SF Model	AMDN Model	ADVAE Model	CIHSA-RTODT	AODC-MDLSV
10	17.66	23.52	20.52	46.63	48.52
20	30.90	46.11	43.41	69.72	71.37
30	42.59	65.00	68.89	92.51	94.86
40	53.51	73.33	78.73	92.55	94.31
50	62.89	83.37	91.89	96.67	98.79
60	71.45	92.69	97.00	98.59	99.81
70	87.44	98.67	98.91	99.03	99.62
80	89.96	94.16	98.07	98.84	99.21
90	91.21	97.71	99.12	98.62	99.89
100	91.56	95.98	97.81	98.22	99.12

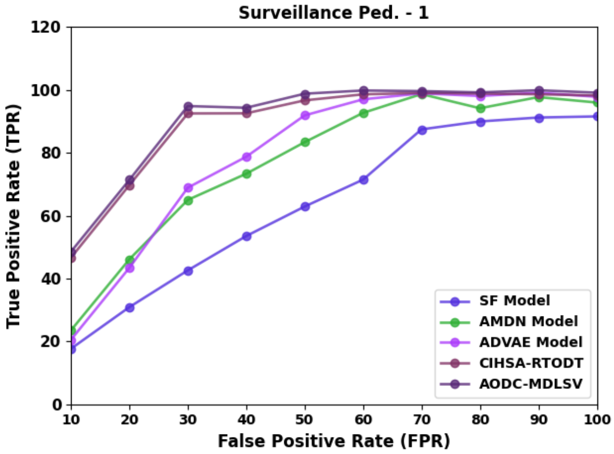

Fig. 6. Comparative analysis of AODC-MDLSV technique on surveillance Ped.-1 dataset

Table 6 and Fig. 7 report the comparative result examination of the AODC-MDLSV methodology with current techniques on surveillance Ped.-2 database. The figure exposed that the AODC-MDLSV model has accomplished maximum ROC values. The outcomes exposed that the SF approach has demonstrated poor outcomes with reduced ROC values. Then, the ADVAE and CIHSA-RTODT model have demonstrated slightly improved ROC values. Likewise, the AMDN model has accomplished reasonable ROC values. Finally, the AODC-MDLSV algorithm has reached maximal performance over the other methods.

TABLE VI

COMPARATIVE OUTCOME OF AODC-MDLSV APPROACH WITH EXISTING SYSTEMS ON SURVEILLANCE PED.-2

ROC	SF Model	AMDN Model	ADVAE Model	CIHSA-RTODT	AODC-MDLSV
10	18.90	27.52	16.53	28.24	56.25
20	27.55	47.67	28.81	60.02	62.01
30	40.76	56.60	68.64	78.80	81.13
40	55.39	74.52	81.60	93.80	96.21
50	73.05	86.90	87.11	96.09	98.98
60	86.18	93.05	95.28	96.55	99.66
70	97.98	98.55	98.77	98.52	99.65
80	98.33	98.48	98.92	99.12	99.93
90	98.54	98.67	98.83	99.24	99.65
100	98.88	98.97	99.11	99.36	99.82

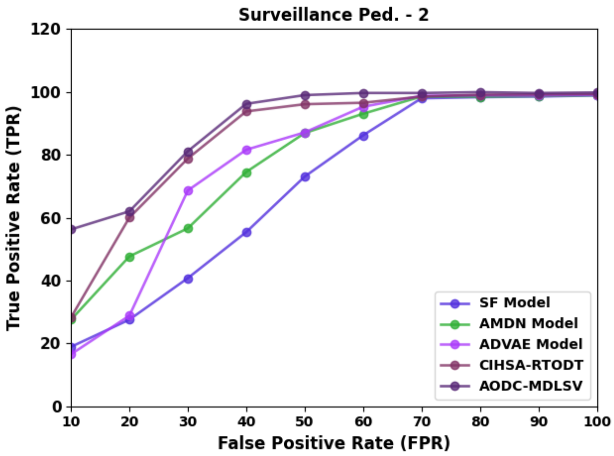

Fig. 7. Comparative analysis of AODC-MDLSV technique on surveillance Ped.-2 dataset

IV. Conclusion

In this study, we have developed a new AODC-MDLSV method for object detection and classification on surveillance videos. The presented AODC-MDLSV technique proficiently detects and classifies the objects into multiple classes. To do so, the AODC-MDLSV technique initially performs object detection using YOLO-v5 method. In the next stage, the AODC-MDLSV technique employed the RVFL model for object classification purposes. Finally, the ABC algorithm is used as a parameter optimization approach to improve detection efficiency. The result analysis of the AODC-MDLSV approach is tested on benchmark video and the outcomes exhibited the superior performance of the AODC-MDLSV method over other existing approaches. In future, hybrid metaheuristics can be derived to improve the detection performance.

REFERENCES

[1] Chandrakar, R., Raja, R., Miri, R., Sinha, U., Kushwaha, A.K.S. and Raja, H., 2022. Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm. *Expert Systems with Applications*, *191*, p.116306.

[2] Luo, X., Wang, Y., Cai, B. and Li, Z., 2021. Moving object detection in traffic surveillance video: new MOD-AT method based on adaptive threshold. *ISPRS International Journal of Geo-Information*, *10*(11), p.742.

[3] Tom, A.J. and George, S.N., 2020. Simultaneous reconstruction and moving object detection from compressive sampled surveillance videos. *IEEE Transactions on Image Processing*, *29*, pp.7590-7602.

[4] Ingle, P.Y. and Kim, Y.G., 2022. Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities. *Sensors*, *22*(10), p.3862.

[5] Kim, J.H., Choi, J.H., Park, Y.H. and Nasridinov, A., 2021. Abnormal situation detection on surveillance video using object detection and action recognition. *Journal of Korea Multimedia Society*, *24*(2), pp.186-198.

[6] Alotaibi, M.F., Omri, M., Abdel-Khalek, S., Khalil, E. and Mansour, R.F., 2022. Computational intelligence-based harmony search algorithm for real-time object detection and tracking in video surveillance systems. *Mathematics*, *10*(5), p.733.

[7] Joy, F. and Vijayakumar, V., 2022. An improved Gaussian Mixture Model with post-processing for multiple object detection in surveillance video analytics. *International journal of electrical and computer engineering systems*, *13*(8), pp.653-660.

[8] Arunnehru, J., 2021. Deep learning-based real-world object detection and improved anomaly detection for surveillance videos. *Materials Today: Proceedings*.

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

[10] Kumar, C. and Punitha, R., 2020, August. Yolov3 and yolov4: Multiple object detection for surveillance applications. In *2020 Third international conference on smart systems and inventive technology (ICSSIT)* (pp. 1316-1321). IEEE.

[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*, pp.611-630.

[13] Mahalingam, T. and Subramoniam, M., 2021. A robust single and multiple moving object detection, tracking and classification. *Applied Computing and Informatics*, *17*(1), pp.2-18.

[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*, 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*, *41*(1), pp.1777-1789.

[16] Katsamenis, Iason, Eleni Eirini Karolou, Agapi Davradou, Eftychios Protopapadakis, Anastasios Doulamis, Nikolaos Doulamis, and Dimitris Kalogeras. "TraCon: A novel dataset for real-time traffic cones detection using deep learning." *In Novel & Intelligent Digital Systems: Proceedings of the 2nd International Conference (NiDS 2022)*, pp. 382-391. Cham: Springer International Publishing, 2022.

[17] Tang, L., Wu, Y. and Yu, L., 2018. A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting. *Applied Soft Computing*, *70*, pp.1097-1108.

[18] Imran, M., Khushnood, R.A. and Fawad, M., 2023. A Hybrid Data-Driven and Metaheuristic Optimization Approach for the Compressive Strength Prediction of High-Performance Concrete. *Case Studies in Construction Materials*, p.e01890.

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

[20] 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.

[21] 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.

[22] 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.