

Mayfly Optimization with Deep Learning-based Robust Object Detection and Classification on Surveillance Videos

Venkatesan Saikrishnan

Department of Computer and Information Science, Faculty of Science, Annamalai University, India
saikrishnan1074@gmail.com (corresponding author)

Mani Karthikeyan

Department of Computer and Information Science, Faculty of Science, Annamalai University, India
karthiaucse@gmail.com

Received: 26 July 2023 | Revised: 16 August 2023 | Accepted: 23 August 2023

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6231>

ABSTRACT

Surveillance videos are recordings captured by video recording devices for monitoring and securing an area or property. These videos are frequently used in applications, involving law enforcement, security systems, retail analytics, and traffic monitoring. Surveillance videos can provide valuable visual information for analyzing patterns, identifying individuals or objects of interest, and detecting and investigating incidents. Object detection and classification on video surveillance involves the usage of computer vision techniques to identify and categorize objects within the video footage. Object detection algorithms are employed to locate and identify objects within each frame. These algorithms use various techniques, namely bounding box regression, Convolutional Neural Networks (CNNs), and feature extraction to detect objects of interest. This study presents the Mayfly Optimization with Deep Learning-based Robust Object Detection and Classification (MFODL-RODC) method on surveillance videos. The main aim of the MFODL-RODC technique lies in the accurate classification and recognition of objects in surveillance videos. To accomplish this, the MFODL-RODC method follows a two-step process, consisting of object detection and object classification. The MFODL-RODC method uses the EfficientDet object detector for the object detection process. Besides, the classification of detected objects takes place using the Variational Autoencoder (VAE) model. The MFO algorithm is employed to enrich the performance of the VAE model. The simulation examination of the MFODL-RODC technique is performed on benchmark datasets. The extensive results accentuated the improved performance of the MFODL-RODC method over other existing algorithms with an output of 98.89%.

Keywords-surveillance videos; object detection; deep learning; classification; computer vision

I. INTRODUCTION

The use of surveillance cameras, also known as Closed-Circuit Television (CCTV), has met a fast development in a global scale. The surveillance systems rely on human supervisors who find some practical activities in video scenes. Monitoring parallel events in surveillance displays is complicated and has constraints [2]. Visual surveillance systems can support detecting and tracking objects with several cameras [3]. An advanced surveillance system includes video and image data acquisition devices, data processing - storage devices, and analysis modules, components which are important for the workflow of the system [4]. Due to the latest advances in Deep Learning (DL) based algorithms, Object Detection (OD) has seen considerable progress during the last few decades [5]. Numerous researchers have explored OD by considering every frame of video as an image for detecting objects in videos. But humans do not observe all the frames as

a self-sufficient image in the process of identifying the objects in videos, instead of tracking a movement concerning a prior frame [6]. Hence, video detection techniques like automotive driving, video surveillance, and intelligent robotics are a combination of object tracking and OD techniques. OD is a standard application in Computer Vision (CV) [7]. The task of identifying objects inside the provided images consists of two wide-ranging parts namely object classification and OD [8], with the latter being far more complex. Furthermore, object classification does not work on images that include multiple objects. The DL techniques are utilized for OD for many reasons [9]. DL algorithms are simpler to use and have improved scalability than traditional ML techniques and their capability for processing the data in their raw forms. Besides, DL algorithms are proficient to learn highly complex features by taking the benefits of various levels of representation [10].

This study presents the Mayfly Optimization with Deep Learning based Robust Object Detection and Classification (MFODL-RODC) method. The main aim of the MFODL-RODC technique lies in the accurate classification and recognition of objects in surveillance videos. To accomplish this, the MFODL-RODC method follows a two-step process: object detection and object classification. The MFODL-RODC method uses the EfficientDet object detector for the object detection process. The classification of detected objects takes place using the Variational Autoencoder (VAE) model. The MFO algorithm is employed to enrich the performance of the VAE model. The simulation validation of the MFODL-RODC technique is performed on benchmark datasets.

II. RELATED WORKS

Authors in [11] suggested a new Computational Intelligence-based HSA for Real-Time OD and Tracking (CIHSART-ODT) method on video surveillance systems. The proposed method focuses largely on tracking and detecting the objects that occur in video clippings. Furthermore, the hyperparameter value of the enhanced RefineDet method is finetuned by implementing the Adagrad optimization model. In addition, an HSA with a Twin SVM (TWSVM) algorithm is used for classifying objects. In [12], a CNN is used for enhancing the effectiveness of OD using a Probabilistic Neural Network (PNN) during the evaluation of the images of surveillance videos. Then, the research makes an effort to analyze the essential theories and principles of perceptual and neural networks that are normally left out. Authors in [13] presented the Multi-Object Detection and Tracking (MODT) method. This technique exploits a Kalman filtering model in order to track the objects shifting in video captions which are converted into morphological operations by employing the region-growing method. Kalman filtering is applied after distinguishing the objects, for parameter optimization by implementing a probability-based grasshopper model. Authors in [14] suggested a precise and fast technique for OD. The transfer learning of an effective pre-trained model to a suitable dataset for its applications was suggested. Fine tuning on these pre-trained models was implemented by running backpropagation and replacing the softmax layer. In [15], the authors introduced a new hybridization of ANN in addition to the Oppositional Gravitational Search Optimization (ANN-OGSO) technique based Moving Vehicle Detection (MVD) model. The presented technique comprises two different stages, i.e. vehicle detection and background generation. Initially, an efficient method was introduced for generating the background. Then, the moving vehicles were detected with the ANN-OGSA algorithm. In [16], a new Background Modeling mechanism was developed by using the Biased Illumination Field FCM algorithm for more accurate detection of moving objects. The non-stationary pixels were separated from the stationary via background elimination. Later, the biased illumination field FCM method was used for improving the segmentation performance by clustering under noise and changing illumination environments. Authors in [17] introduced a new architecture which integrates enhanced GMM with postprocessing for the OD process. In this work, pre-processing and post-processing of videos were taken as extrinsic improvements. GMM with parameter initialization was taken

as an essential enhancement. The incorporation of the morphological function with GMM assists segmentation and enhances detection accuracy by decreasing the false positives.

III. THE PROPOSED MODEL

In this study, the novel MFODL-RODC approach for effective object recognition in surveillance videos is introduced. The purpose of the MFODL-RODC approach lies in the accurate classification and detection of objects in surveillance videos. To accomplish this, the MFODL-RODC method follows a two-step process consisting of the EfficientDet-based object detection and MFO with VAE-based object classification. Figure 1 demonstrates the overall flow of the MFODL-RODC algorithm. The suggested technique is put under simulation by employing Python 3.6.5 tool on a PC i5-8600k, 250GB SSD, GeForce 1050Ti 4GB, 16GB RAM, and 1TB HDD. The parameter set up is: learning rate: 0.01, activation: ReLU, epoch count: 50, dropout: 0.5, and batch size: 5.

A. Stage I: Object Detection

In the initial phase, the EfficientDet architecture is utilized. EfficientDet is an OD model that uses optimization and backbone tweaks, such as the use of BiFPN, and a compound scale model that homogeneously scales the width, depth, and resolution of the feature network and class or box prediction network simultaneously [18]. First, the EfficientDet is exploited as a backbone network, later the recurrent BiFPN acts as a feature extraction model for making multiscale feature fusion of P3-P7 features from the EfficientNet. Lastly, the fused features are fed into the box and class predictive network for classification and bounding box prediction. The input can be taken as a low-dimensional compressed representation that is extended to a high dimension by a 1×1 convolutional layer. The feature is filtered with the depth-wise separable convolution for spatial data encoding.

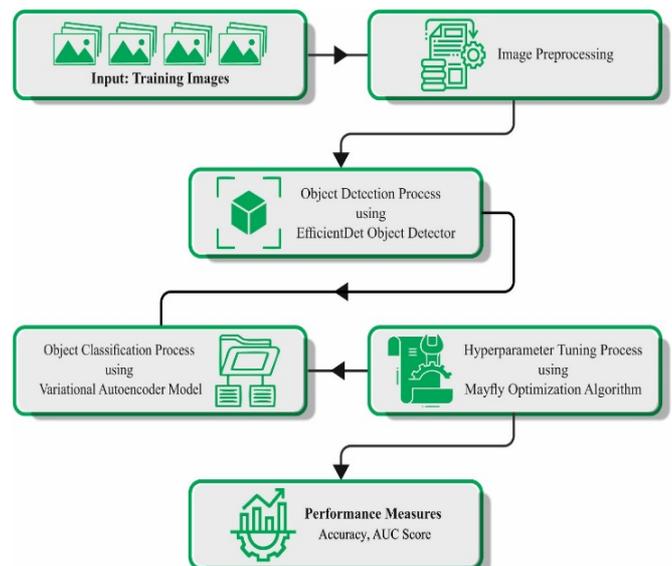


Fig. 1. Overall flow of the MFODL-RODC model.

The SE block employs average pooling for squeezing the spatial data. The aggregated features are set to excitation operators for fully capturing channel-wise dependency. The output features are regarded as a weight of all the channels. Finally, the features are connected to the initial inputs with a reversed residual model for weighting the channels. BiFPN makes various optimizations on Path Aggregation Network (PANet) for further fusing high-level factors in the simplify bi-directional network. Furthermore, all the inputs are added with a further weight such that the network can learn the feature involvement at multiple resolution scales. The final BiFPN enables effective multiscale feature fusion by incorporating weighted feature fusion and bi-directional cross-scale connection. The compound scaling method is proposed due to its accuracy and efficiency. Depth, width, and resolution of the images are scaled-up together by a Φ coefficient, whereby the resource used can be calculated.

B. Object Classification

Once the objects are detected, the process of classification is achieved by employing the VAE technique. A VAE comprises of the encoded network $f_{\phi}^{enc} : \mathcal{X} \rightarrow \mathcal{Z}$ and the decoded network $f_{\theta}^{dec} : \mathcal{Z} \rightarrow \mathcal{X}$, where \mathcal{X} implies the input variable (object) and \mathcal{Z} represents the hidden object domain z [19]. In VAE, input objects x and hidden parameters z are assumed to have arbitrary variables. Let's assume that $p_{\theta}(\mathcal{X})$ and $p_{\theta}(\mathcal{Z})$ are the x and z 's probability functions, correspondingly, where θ denotes the group of parameters for illustrating the probability. The encoded network f_{ϕ}^{enc} is expressed as a conditional probability $q_{\phi}(\mathcal{Z}|\mathcal{X})$ that is assumed as an approximation of $p_{\theta}(\mathcal{Z}|\mathcal{X})$. The decoded network f_{θ}^{dec} and ϕ are expressed as the conditional probability $p_{\theta}(\mathcal{X}|\mathcal{Z})$ and the encoded parameters:

$$J_{VAE}(\phi, \theta) = E_{z \sim q_{\phi}(z|x)} \log p_{\theta}(x|z) - \beta D_{KL}(q_{\phi}(z|x) || p_{\theta}(z)) \quad (1)$$

In VAE, the input or the hidden parameters are assumed as having arbitrary variables. The purpose of VAE training is to maximize a main function which comprises 2 terms: (i) Kullback-Leibler (KL) divergence, (ii) reconstruction. The first term inspires the encoding for producing hidden objects that follow past distribution, and the second term makes sure that the input objects are reconstructed in the hidden variable. The VAE main function decreases to objective functions of the typical VAEs once the hyperparameter β is fixed to 1. VAEs with $\beta \neq 1$ are termed as β -VAEs. Figure 2 represents the infrastructure of VAE.

To improve the detection rate of the VAE model, the MFO algorithm is employed. MFO is a robust hybrid optimization technique inspired by the behaviors of MFS in mating. It implements and enhances the global searching of PSO [20]. This technique disregards the lifetime of MFs and in its place considers an adult directly after hatching while only the stronger one survives. A set of male and as female MFS are arbitrarily created. The location vector $P = [p_1, p_2, \dots, p_{d_{Max}}]^T$ signifies the search space, and the agents that perform the search are seeded at first. The main function (OF) estimates the efficiency of the location vector (x). With the velocity vector,

the MFs location is studied considering the revised movement path that can be informed by the individual and social movement experience $K = [k_1, k_2, \dots, k_{d_{Max}}]^T$. Based on its present optimal location, MF moves up or down the search graph.

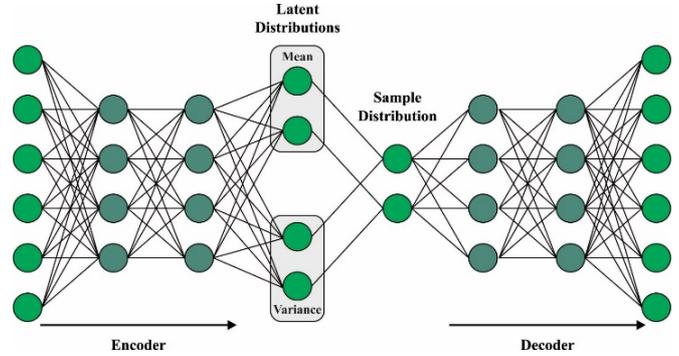


Fig. 2. Framework of VAE.

A male MF's status can be revised by:

$$p_m(t + 1) = k_m(t + 1) + p_m(t) \quad (2)$$

For the i^{th} MF, $p_m(t)$ denotes the existing position and $p_m(0)$ falls among x_{Min} and x_{Max} . The subsequent time step's MFs position and velocity are denoted by $p_m(t + 1)$ and $k_m(t + 1)$, respectively.

The constant speed is evaluated with the male MFs nuptial dance occurring in a certain height.

$$k_{md}(t + 1) = v_{md}(t) + q_1 \times \exp(-\zeta^{D_p^2}) \times (pbest_{md} - p_{md}(t)) + q_2 \times \exp(-\zeta^{D_g^2}) \times (gbest_d - p_{md}(t)) \quad (3)$$

where q_1 and q_2 indicate the attractive constants that determine the comparative significance of the mental and social elements. MFs cannot see one another very well once they are in a σ environment. Based on (5) and (6), we could define the D_p and D_g distances that p_i has with $pbest_m$ and $gbest$, correspondingly. The i^{th} agent velocity from the d^{th} dimension is represented as k_{md} , while its location can be represented as p_{md} . d shows the dimension index ranges from 1 to d_{Max} , which is the maximum number of dimensions. The optimum location $pbest$ is apprehended by the i^{th} agent at the d^{th} dimension and is evaluated by (4). The quality-defining OF for these solutions is shown as $f(\cdot)$:

$$pbest_m = \begin{cases} x_m(t + 1), f(x_m(t + 1)) < f(pbest_m) \\ pbest_m, f(x_m(t + 1)) \geq f(pbest_m) \end{cases} \quad (4)$$

$$D_p^2 = (\sum_{n=1}^{d_{Max}} (p_{md} - pbest_m))^0.5 \quad (5)$$

$$D_g^2 = (\sum_{d=1}^{d_{Max}} (p_{md} - gbest))^0.5 \quad (6)$$

$$k_{md}(t + 1) = k_{md}(t) + ND \times \omega \quad (7)$$

In (7), ND denotes the nuptial dance coefficient and ω indicates an arbitrary integer within $[-1, 1]$.

Female MFs do not swarm like males do. Instead, they head straight to the males for mating. Utilizing $r_m(t)$, it is understood that the i^{th} female MF is placed in the searching space after utilizing (8) for adjusting place:

$$r_m(t + 1) = k_m(t + 1) + r_m(t) \tag{8}$$

For modeling this phenomenon, it can be considered that one attractive female can be drawn to one attractive males, after another attractive female is attracted to the next most attractive male, etc. The below equation depicts the speed:

$$k_{md}(t + 1) = \begin{cases} k_{md}(t) + q_2 \times \exp(-\zeta_f^{D_f^2}) \times (p_{md}(t) - r_{md}(t)) \\ f(r_m) > f(p_m) \\ u_{md}(t) + q_w \times \omega, f(r_m) \leq f(p_m) \end{cases} \tag{9}$$

$r_{md}(t)$ and $u_{md}(t)$ define the place and velocity of the i^{th} female MF from the d^{th} dimension at time t , correspondingly. Male and female MFs separate distances are represented by D_{if}^2 . The coefficient of walking, q_w , is selected arbitrarily.

The crossover operator is utilized for modeling the MF mating performance defined as one male and one female can be chosen in all the sets as parents, while males can be attracted to particular females. The selection depends on both chance and the main function. With the employment of the subsequent formulas, the offspring of crossover is predicted:

$$\alpha 1 = \beta \times male + (1 - \beta) \times female \tag{10}$$

$$\alpha 2 = \beta \times female + (1 - \beta) \times male \tag{11}$$

The first two generations of this family are represented by $\alpha 1$ and $\alpha 2$. β'' refers to an arbitrary number in a certain interval. *Male* and *female* signify the biological parents. Fitness choice is a vital feature of the MFO system. An encoding solution was employed to process the better candidate result. Presently, the value of accuracy is the major condition employed to plan a FF.

$$Fitness = \max(P) \tag{12}$$

$$P = \frac{TP}{TP+FP} \tag{13}$$

FP and TP represent the false and true positive values.

IV. RESULTS AND DISCUSSION

In this section, the object recognition and classification outputs of the MFODL-RODC technique are tested on the UCSDPed2 dataset [21] comprising of 360 frames as portrayed in Table I.

TABLE I. DATASET DESCRIPTION

Dataset	Test dataset	Frame number	Time (s)
UCSDped2	Pedestrian1	360	12
	Pedestrian2		

Figures 3 and 4 depict instances, and the initial and recognized images. Table II represents the average accuracy study of the MFODL-RODC method with other approaches [11, 22-24]. The outcomes show that the MFODL-RODC method has improved performance in comparison with object

detectors with maximum accuracy of 98.89% and 96.23% on SPed-1 and SPed-2 datasets, respectively. The comparative AUC results of the MFODL-RODC technique on two datasets are given in Table III. The outputs imply that the MP-PCA, SF, and SFMP-PCA techniques gave worse results with minimal AUC values.

TABLE II. COMPARISON OF THE AVERAGE ACCURACY OF MFODL-RODC WITH OTHER APPROACHES

Methods	MFODL-RODC	CIHSART-ODT	DLADT	Region CNN	FR-CNN
Surveillance Ped. - 1	98.89	98.00	97.00	97.00	85.00
Surveillance Ped. - 2	96.23	91.00	90.00	87.00	82.00

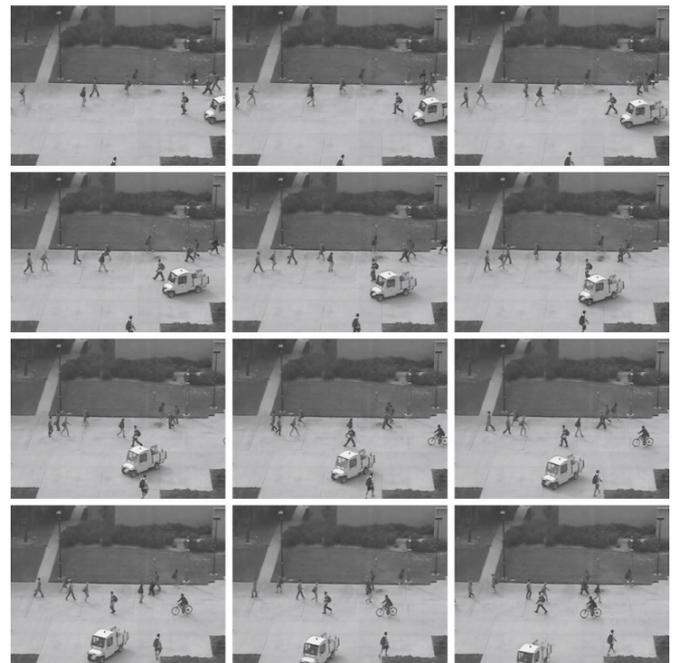


Fig. 3. Sample images.



Fig. 4. (a) Original and (b) detected images.

On the other hand, the MDT method exhibited a moderate AUC value. A-MDN, AD-VAE, and CIHSART-ODT methods showed reasonable AUC values. But the MFODL-RODC method showed greater performance with the highest AUC values of 99.28% and 96.07% on SPed-1 and SPed-2 datasets.

TABLE III. AUC OUTPUT COMPARISON

Model	SPed-1	SPed-2
MP-PCA	61.01	69.92
SF	66.74	55.96
SFMP-PCA	67.25	61.33
MDT	82.05	82.99
A-MDN	91.71	91.25
AD-VAE	95.39	92.47
CIHSART-ODT	97.12	93.92
MFODL-RODC	99.28	96.07

In Table IV, the Running Time (RT) comparison of the MFODL-RODC technique with recent models on two datasets is shown. The results illustrate that the MDT and SCLF models have poor performance with maximum RT values. The A-MDN model gave slightly decreased RT. Although the AD-VAE and CIHSART-ODT models have obtained close RT values, the MFODL-RODC technique ensured its supremacy with minimal RT of 1.93s and 2.20s on pedestrian-1 and pedestrian-2 datasets, respectively.

TABLE IV. RT OUTPUT COMPARISON

Model	SPed-1	SPed-2
MDT	20.53	22.88
SCLF	20.01	18.54
A-MDN	11.83	13.04
AD-VAE	3.94	6.09
CIHSART-ODT	2.57	4.04
MFODL-RODC	1.93	2.20

TABLE V. ROC OUTPUT COMPARISON ON SPED-1

ROC	SF	A-MDN	AD-VAE	CIHSART-ODT	MFODL-RODC
10	16.65	24.63	20.99	47.51	49.01
20	31.14	46.03	44.43	70.80	72.85
30	41.86	64.90	67.52	91.16	96.24
40	52.58	74.63	80.31	93.35	94.63
50	61.65	82.55	92.69	95.77	99.29
60	70.51	92.12	95.93	98.58	99.89
70	87.53	99.12	100.40	98.06	99.78
80	88.51	93.43	97.86	98.08	99.65
90	89.89	99.25	97.86	99.69	99.91
100	90.20	95.27	96.76	96.87	99.48

TABLE VI. ROC OUTPUT COMPARISON ON SPED-2

ROC	SF	A-MDN	AD-VAE	CIHSART-ODT	MFODL-RODC
10	19.27	26.64	18.30	28.16	56.07
20	28.75	48.29	29.00	61.50	62.56
30	41.05	57.25	69.70	80.01	82.52
40	55.44	74.39	82.33	93.32	97.16
50	74.39	88.12	88.15	97.69	99.18
60	87.87	93.08	95.35	97.18	99.75
70	98.40	99.06	99.02	97.14	99.70
80	98.65	99.71	99.09	98.59	99.96
90	98.06	98.71	98.42	99.01	99.79
100	98.51	97.79	98.52	100.83	99.89

Tables V and VI shows the comparison of the ROC outputs on the SPed-1 and SPed-2 databases. In SPed-1, the SF and AD-VAE techniques showed poor detection performance, the A-MDN model has slightly improved results, and the CIHSART-ODT model has obtained considerably enhanced

performance. Nevertheless, the MFODL-RODC technique showed efficient output with maximum ROC values. In SPed-2 the SF and AD-VAE approach demonstrated worse detection performance, the A-MDN model resulted in somewhat improved outcome, and the CIHSART-ODT model obtained considerably better performance. However, the MFODL-RODC approach has illustrated effectual results with maximal ROC values. These outcomes ensured the improved accomplishment of the MFODL-RODC approach.

V. CONCLUSION

The novel MFODL-RODC approach for effective object recognition in surveillance videos is introduced in this paper. The purpose of the study lies on the accurate classification and detection of objects in surveillance videos. To accomplish this, the MFODL-RODC technique follows a two-step process, consisting of EfficientDet-based object detection and MFO with VAE-based object classification. The MFO technique is employed to improve the solution of the VAE approach. The simulation validations of the MFODL-RODC approach was conducted on a benchmark database. The results highlighted the enhanced outcome of the MFODL-RODC methodology over other known techniques. In the future, advanced DL models will be used to classify the detected objects. In addition, the computational complexity of the proposed model needs to be examined on large-scale real-time datasets.

REFERENCES

- [1] P. Y. Ingle and Y.-G. Kim, "Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities," *Sensors*, vol. 22, no. 10, Jan. 2022, Art. no. 3862, <https://doi.org/10.3390/s22103862>.
- [2] R. K. Tripathi, A. S. Jalal, and S. C. Agrawal, "Abandoned or removed object detection from visual surveillance: a review," *Multimedia Tools and Applications*, vol. 78, no. 6, pp. 7585–7620, Mar. 2019, <https://doi.org/10.1007/s11042-018-6472-9>.
- [3] C. Fathy and S. N. Saleh, "Integrating Deep Learning-Based IoT and Fog Computing with Software-Defined Networking for Detecting Weapons in Video Surveillance Systems," *Sensors*, vol. 22, no. 14, Jan. 2022, Art. no. 5075, <https://doi.org/10.3390/s22145075>.
- [4] Mohana and H. R. Aradhya, "Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 12, Jun. 2019, <https://doi.org/10.14569/IJACSA.2019.0101269>.
- [5] N. Funde, P. Paranjape, K. Ram, P. Magde, and M. Dhabu, "Object Detection and Tracking Approaches for Video Surveillance Over Camera Network," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Coimbatore, India, Mar. 2019, pp. 1171–1176, <https://doi.org/10.1109/ICACCS.2019.8728518>.
- [6] S. Khan and L. AlSuwaidan, "Agricultural monitoring system in video surveillance object detection using feature extraction and classification by deep learning techniques," *Computers and Electrical Engineering*, vol. 102, Sep. 2022, Art. no. 108201, <https://doi.org/10.1016/j.compeleceng.2022.108201>.
- [7] F. Pérez-Hernández, S. Tabik, A. Lamas, R. Olmos, H. Fujita, and F. Herrera, "Object Detection Binary Classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance," *Knowledge-Based Systems*, vol. 194, Apr. 2020, Art. no. 105590, <https://doi.org/10.1016/j.knsys.2020.105590>.
- [8] S. Nuanmeesri, "A Hybrid Deep Learning and Optimized Machine Learning Approach for Rose Leaf Disease Classification," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7678–7683, Oct. 2021, <https://doi.org/10.48084/etasr.4455>.

- [9] L. Loyani and D. Machuve, "A Deep Learning-based Mobile Application for Segmenting Tuta Absoluta's Damage on Tomato Plants," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7730–7737, Oct. 2021, <https://doi.org/10.48084/etasr.4355>.
- [10] M. V. Daithankar and S. D. Ruikar, "Analysis of the Wavelet Domain Filtering Approach for Video Super-Resolution," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7477–7482, Aug. 2021, <https://doi.org/10.48084/etasr.4262>.
- [11] M. F. Alotaibi, M. Omri, S. Abdel-Khalek, E. Khalil, and R. F. Mansour, "Computational Intelligence-Based Harmony Search Algorithm for Real-Time Object Detection and Tracking in Video Surveillance Systems," *Mathematics*, vol. 10, no. 5, Jan. 2022, Art. no. 733, <https://doi.org/10.3390/math10050733>.
- [12] B. Dhiyanesh, K. Rajesh Kanna, S. Rajkumar, and R. Radha, "Improved Object Detection in Video Surveillance Using Deep Convolutional Neural Network Learning," in *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, Aug. 2021, <https://doi.org/10.1109/I-SMAC52330.2021.9640894>.
- [13] M. Elhoseny, "Multi-object Detection and Tracking (MODT) Machine Learning Model for Real-Time Video Surveillance Systems," *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, pp. 611–630, Feb. 2020, <https://doi.org/10.1007/s00034-019-01234-7>.
- [14] M. Ahmadi, W. Ouarda, and A. M. Alimi, "Efficient and Fast Objects Detection Technique for Intelligent Video Surveillance Using Transfer Learning and Fine-Tuning," *Arabian Journal for Science and Engineering*, vol. 45, no. 3, pp. 1421–1433, Mar. 2020, <https://doi.org/10.1007/s13369-019-03969-6>.
- [15] A. Appathurai, R. Sundarasekar, C. Raja, E. J. Alex, C. A. Palagan, and A. Nithya, "An Efficient Optimal Neural Network-Based Moving Vehicle Detection in Traffic Video Surveillance System," *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, pp. 734–756, Feb. 2020, <https://doi.org/10.1007/s00034-019-01224-9>.
- [16] S. Kalli, T. Suresh, A. Prasanth, T. Muthumanickam, and K. Mohanram, "An effective motion object detection using adaptive background modeling mechanism in video surveillance system," *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 1, pp. 1777–1789, Jan. 2021, <https://doi.org/10.3233/JIFS-210563>.
- [17] F. Joy and D. V. Vijayakumar, "An improved Gaussian Mixture Model with post-processing for multiple object detection in surveillance video analytics," *International Journal of Electrical and Computer Engineering Systems*, vol. 13, no. 8, pp. 653–660, Oct. 2022, <https://doi.org/10.32985/ijeces.13.8.5>.
- [18] Y. Ye *et al.*, "An Adaptive Attention Fusion Mechanism Convolutional Network for Object Detection in Remote Sensing Images," *Remote Sensing*, vol. 14, no. 3, Jan. 2022, Art. no. 516, <https://doi.org/10.3390/rs14030516>.
- [19] O. Boyar, K. Iwata, H. Hanada, and I. Takeuchi, "Sample-Efficient De Novo Chemical Design with Latent Reconstruction-Aware Variational Autoencoder," presented at the 37th Annual Conference of the Japanese Society for Artificial Intelligence, Apr. 2023.
- [20] N. Alqahtani *et al.*, "Deep Belief Networks (DBN) with IoT-Based Alzheimer's Disease Detection and Classification," *Applied Sciences*, vol. 13, no. 13, Jan. 2023, Art. no. 7833, <https://doi.org/10.3390/app13137833>.
- [21] "UCSD Anomaly Detection Dataset." <http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm>.
- [22] I. V. Pustokhina, D. A. Pustokhin, T. Vaiyapuri, D. Gupta, S. Kumar, and K. Shankar, "An automated deep learning based anomaly detection in pedestrian walkways for vulnerable road users safety," *Safety Science*, vol. 142, Oct. 2021, Art. no. 105356, <https://doi.org/10.1016/j.ssci.2021.105356>.
- [23] M. Xu, X. Yu, D. Chen, C. Wu, and Y. Jiang, "An Efficient Anomaly Detection System for Crowded Scenes Using Variational Autoencoders," *Applied Sciences*, vol. 9, no. 16, Jan. 2019, Art. no. 3337, <https://doi.org/10.3390/app9163337>.
- [24] B. S. Murugan, M. Elhoseny, K. Shankar, and J. Uthayakumar, "Region-based scalable smart system for anomaly detection in pedestrian walkways," *Computers & Electrical Engineering*, vol. 75, pp. 146–160, May 2019, <https://doi.org/10.1016/j.compeleceng.2019.02.017>.