



Segmentation Based Real Time Anomaly Detection and Tracking Model for Pedestrian Walkways

B. Sophia^{1,*} and D. Chitra²

¹Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, Tamilnadu, India

²Department of Computer Science and Engineering, P.A College of Engineering and Technology, Pollachi, Tamilnadu, India

*Corresponding Author: B. Sophia. Email: sophiabmecse@gmail.com

Received: 11 March 2022; Accepted: 17 April 2022

Abstract: Presently, video surveillance is commonly employed to ensure security in public places such as traffic signals, malls, railway stations, etc. A major challenge in video surveillance is the identification of anomalies that exist in it such as crimes, thefts, and so on. Besides, the anomaly detection in pedestrian walkways has gained significant attention among the computer vision communities to enhance pedestrian safety. The recent advances of Deep Learning (DL) models have received considerable attention in different processes such as object detection, image classification, etc. In this aspect, this article designs a new Panoptic Feature Pyramid Network based Anomaly Detection and Tracking (PFPN-ADT) model for pedestrian walkways. The proposed model majorly aims to the recognition and classification of different anomalies present in the pedestrian walkway like vehicles, skaters, etc. The proposed model involves panoptic segmentation model, called Panoptic Feature Pyramid Network (PFPN) is employed for the object recognition process. For object classification, Compact Bat Algorithm (CBA) with Stacked Auto Encoder (SAE) is applied for the classification of recognized objects. For ensuring the enhanced results better anomaly detection performance of the PFPN-ADT technique, a comparison study is made using University of California San Diego (UCSD) Anomaly data and other benchmark datasets (such as Cityscapes, ADE20K, COCO), and the outcomes are compared with the Mask Recurrent Convolutional Neural Network (RCNN) and Faster Convolutional Neural Network (CNN) models. The simulation outcome demonstrated the enhanced performance of the PFPN-ADT technique over the other methods.

Keywords: Panoptic segmentation; object detection; deep learning; tracking model; anomaly detection; pedestrian walkway

1 Introduction

Currently, surveillance camera is mounted in public places such as traffic signals, roads, shopping malls, railway stations, banks, etc. to improve security. But it is a difficult task to continuously monitor the video at a faster rate [1]. The result is ineffective use of surveillance camera and require presence of humans to



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

monitor. The major challenge in video monitoring is the recognition of anomalies such as thefts, accidents, crimes, or other illegal activities. The anomaly action doesn't appear more frequently than regular events [2]. Hence, to prevent human resources and time from being wasted, smart computer vision algorithm has been proposed for automatically detecting anomalies at a fast rate [3]. The main objective of anomaly detection method is to raise an alert at situations where the event typically diverges from the actual events. Therefore, the anomaly detection is regarded as a coarse-level video understanding issue that recognizes the existence of anomalies from frequent events [4].

Once the anomaly is detected effectively, it is separated into one of the certain actions using classification algorithm. The anomaly detection method is enhanced for identifying certain anomalies, *e. g.*, traffic accident detector and violence detector. At the same time, it must be noted that this solution shouldn't be discriminated against to recognize other anomalies [5]. Some real time anomalies are proved to be unique and complex, while it is hard to identify all the anomalies. Hence, it is preferred to propose the anomaly detection technique that doesn't depend on accessible data about the anomaly. In recent times, distinct approaches have been deployed to compute the pedestrian prediction that suits bounding box for a pedestrian existing in an image [6]. It has received maximal interest from the designers of computer vision and the important component for different human-based areas including automated traffic signaling, person examination, driverless cars, and so on. But the predetermined model is not fit to resolve the difficulty of algorithm called scaling problem which remains unchanged and causes the result of pedestrian detection method. The conventional method has been handled to resolve the scaling problem on the two-dimensional scale [7]. Firstly, brute-force data is increased for improving the capacity of scale-invariance method. After that, a single model using many scale filters has been used in each sample with distinct sizes. But the existence of intra-class variance of maximal and tiny samples is complex to address the individual approaches and diverse feature responses [8]. To utilize dramatically varying attributes with diverse scales, the divide-and-conquer method is used [9] to resolve the complex scale variance problem. Eventually, Deep Learning (DL) depending on anomaly prediction method is deployed. At first, Convolutional Neural Network (CNN) has been applied and classified the existence of object. It has faced some problems such as aspect ratios of objects and massive spatial locations from an image [10]. This article designs a new Panoptic Feature Pyramid Network based Anomaly Detection and Tracking (PFPN-ADT) model for pedestrian walkways. The proposed model majorly aims to the recognition and classification different anomalies present in the pedestrian walkway like vehicles and skaters, etc. The proposed model involves panoptic segmentation model, called Panoptic Feature Pyramid Network (PFPN) is employed for the object recognition process. For object classification, Compact Bat Algorithm (CBA) with Stacked Auto Encoder (SAE) is applied for the classification of recognized objects. For ensuring the enhanced results better anomaly detection performance of the PFPN-ADT technique, a comparison study is made using University of California San Diego (UCSD) Anomaly data and other benchmark datasets.

2 Literature Review

Ye et al. [11] introduced a method for detecting and tracking Unmanned Aerial Vehicles (UAVs) from an individual camera installed on distinct UAVs. At first, background motion is estimated through a perception transformation method and later recognize moving object candidates in the background subtracted image via Deep Learning classifiers trained on automatically labelled datasets. For every moving object candidate, the study discovers Spatio-temporal trait via optical flow matching and later prune them according to motion pattern than the background. Kalman filter is employed on pruned moving objects to enhance temporal dependency amongst the candidate recognitions. Hsu et al. [12] presented a passenger flow counting method for the bus. Firstly, a direct method is designed for understanding the open state of the door. Then, a single shot multi-box detector is utilized for learning the feature of passengers and distinguishing

them. At last, a particle filter using a three-phase cascaded data association arrangement is utilized for tracking passengers. Wang et al. [13] developed an architecture called Moving-object Proposal Generation and Prediction framework (MPGP) for reducing the search space and generating accurate proposals that could decrease computation cost. Additionally, examine the relationship of moving region in feature map of distinct layers and forecast candidate based on the result of preceding frame. In [14], the moving object detection using TensorFlow object detection Application Programming Interface (API). Furthermore, position of the detected object is passed to the object tracking method. A new CNN based object tracking approach has been utilized for strong object detection. The presented method is capable of detecting the object in distinct occlusion and illumination. The authors in [15] proposed an approach to detect, count, and track vehicles in roundabout videos. There are two major contributions, (i) the moving vehicle is taken into account for tracking, and (ii) the vehicle truck output through the object tracking algorithm is treated for reducing the false track rate. The vehicle detection can be implemented by utilizing YOLOv4, and vehicle tracking all over the video can be attained by Kalman filter or DeepSORT approach. In [16], a region-based detection method is presented for detecting the bare hand-like region using movement information and skin color. For tracking consecutive frames, Detection and Tracking (DaT) approach has been presented. A smoothen trajectory is employed to train a separate Deep Neural Network (DNN) for recognizing sixty isolated dynamic gestures.

3 The Proposed Model

In this study, a new PFPN-ADT system has been developed for the recognition and classification of different anomalies present in the pedestrian walkway like vehicles and skaters, etc. The PFPN-ADT model encompasses PFPN based object detection, Stacked Auto Encoder (SAE) based object classification, and (Compact Bat Algorithm) based parameter optimization.

3.1 Object Detection Module: PFPN Model

Initially, the surveillance video is converted into a set of frames and object is identified at every frame using the PFPN model [17,18]. Panoptic FPN is a single, simple network baseline that aims at achieving joint tasks, topmost efficiency on instance and semantic segmentation: panoptic segmentation [17]. It begins from Mask Region-Based Convolutional Neural Network (RCNN) using Feature Pyramid Network (FPN), a stronger instance segmentation baseline, and makes slight variations to create a semantic segmentation dense-pixel output.

FPN: Starts with brief analysis of FPN. It takes a typical network with features at spatial resolutions (for example Inception V3). The top-down pathway begins from the deep layer network and gradually up samples when adding in converted version of high-resolution feature from the bottom-up pathways. FPN generate a pyramid, generally with scales from 1/32 to 1/4 resolution, while all the pyramid levels have similar channel dimension (256 by default).

Baseline network: The inception model is a kind of DNN structure designed by an author called Szegedy et al. originally in 2014. The structure of inception model and the traditional CNN method are distinct from one another. So the inception model is inception block that implies lapping the similar input tensor with filter and concatenating the result. There are different versions of the inception model. Inception-V3 is an enhanced version of earlier inception model that is Inception-V1 and Inception-V2. Inception-V3 comprises overall 24M variables. Fig. 1 illustrates the framework of Inception V3 Model.

Instance segmentation branch: The objective of FPN, and especially the usage of similar channel dimensions for each pyramid level, makes it easier for attaching region-based object detectors such as Fast RCNN [18]. Fast RCNN implements Region of Interest (RoI) pooling on distinct pyramid levels and employs a shared network branch for predicting a class label and the refined box for all the regions.

For output instance segmentation, apply Mask RCNN that extends Fast RCNN by including FCN to forecast a binary segmentation mask for all the candidate regions.

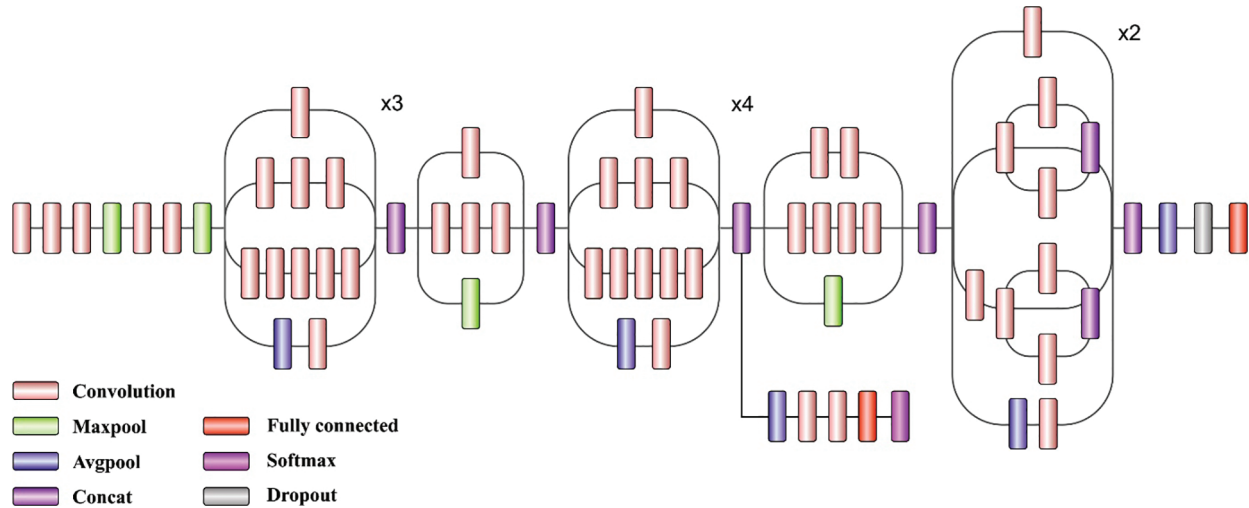


Figure 1: Structure of inception V3 model

Panoptic FPN: Its goal is to alter Mask RCNN using FPN to permit pixel-wise semantic segmentation calculation. But to accomplish precise prediction, the feature utilized for the task must: (1) have appropriately higher resolution for capturing fine structure, (2) sufficiently encode rich semantics to precisely forecast class label, and (3) capture multiscale data to forecast stuff region at various resolutions. Even though FPN was proposed for detecting objects, this requirement—higher-resolution, rich, multiscale feature—recognizes precisely the features of FPN.

Semantic segmentation branch: For generating the output from the FPN feature, a simple model is presented to combine the data from each level of the FPN pyramid into individual output. Starting from the deep FPN level (at 1/32 scale), three upsampling phases are implemented to produce a feature map at 1/4 scale, while all the up-sampling stages comprise of group norm, 3×3 convolution, $2 \times$ bilinear upsampling, and ReLU. A last 1×1 convolutional, $4 \times$ bilinear upsampling, and softmax are utilized for generating the per pixel class label at original image resolution. Besides stuff classes, this branch outputs an ‘other’ class for each pixel that belongs to object.

3.2 Object Classification Module: SAE Model

Next to the detection of objects, the classification process is performed by the use of SAE model [19]. AE is a type of unsupervised learning infrastructure which maintains 3 layers such as input, hidden, and output layers. The procedure of AE trained has 2 parts encoding and decoding [19]. The encoder was utilized to map an input dataset as to hidden representations, and decoding was mentioned that recreating input data in the hidden demonstration. To provide the unlabelled input data $\{x_n\}_{n=1}^N$, whereas $x_n \in \mathbb{R}^{m \times 1}$, h_n implies the hidden encoding vectors computed in x_n , and \hat{x}_n signifies the decoded vectors of resultant layer. Therefore, the encoding procedure is as follows:

$$h_n = f(W_1 x_n + b_1) \quad (1)$$

where f implies the encoded functions, W_1 stands for the weighted matrix of encoding, and b_1 refers to the bias vector. The decoding procedure was determined as:

$$\hat{x}_n = g(W_2 h_n + b_2) \tag{2}$$

where g signifies the decoded functions, W_2 represents the weighted matrix of decoding, and b_2 stands for the bias vectors. The variable set of AE model is the optimization for minimizing the reform error:

$$\mathcal{O}(\Theta) = \arg \min_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L(x^i, \hat{x}^i) \tag{3}$$

In which L denotes the loss function $L(x, \hat{x}) = \|x - \hat{x}\|^2$. The infrastructure of SAE is stack n AEs as to n hidden states with unsupervised layer-wise learning technique and afterward fine-tuning with supervised technique. Thus, the SAE based approach was separated into 3 phases:

- i) Trained the primary AE by input dataset and attain the learned feature vectors;
- ii) The feature vectors of past layer were utilized as input to the subsequent layer, and the process was repeating still the trained ends.
- iii) Afterward, every hidden layer is trained, BP technique was utilized for minimizing the cost functions and upgrading the weight with labeled trained set for achieving fine-tuned.

3.3 Parameter Optimization Using CBA

For optimally modifying the parameters involved in the SAE model, the CBA can be applied to it. The purpose of the compact technique is for simulating the functions of population-based technique of BA [20] under the version with significantly lower saved variable memory. The real population of solution of BA was changed as to compact technique with creating a distributed data system, such as Perturbation Vector(PV). PV refers to the probabilistic method to population of solution.

$$PV^t = [\mu^t, \delta^t], \tag{4}$$

where δ and μ represent 2 variables of mean of vectors PV and Standard Deviation (SD), as t represents the present time [21]. The amplitude of PDFs is normalization with keep their region equivalent to 1 since by attaining around appropriate incorrect it can be uniform distribution with complete shape. The real valued prototype vector was utilized for maintaining sampling probabilistic to arbitrarily create modules of candidate solutions. This vector function was distributed dependent upon the Estimated Distribution Algorithm (EDA). As some novel creating candidate is storing from the memory it could not all overpopulation of solutions saved from memory.

The possibility is that the evaluated distribution is trends, driving novel candidate forwarding to Fast Forward (FF). The candidate solution is created probabilistically in the vector, and modules from the optimum solutions were utilized for making small modifications to probability from the vector [21]. The candidate solution χ_i equivalent to the place of virtual bats are created by (μ_i, δ_i) .

$$P(x) = \frac{\frac{(x - \mu_i)^2}{2\delta_i^2}}{\delta_i \left(\operatorname{erf} \left(\frac{\mu_i + 1}{\sqrt{2}\delta_i} \right) - \operatorname{erf} \left(\frac{\mu_i - 1}{\sqrt{2}\delta_i} \right) \right)} \tag{5}$$

where $P(x)$ refers to the likelihood distribution of PV which expressed the truncated Gaussian PDF connected to μ and δ . A novel candidate solution was created by being iteratively biased near a promising region of better solutions. It is attained that all components of probability vectors with learning the earlier generation. Also, the CDF requires the distribution of multi-variate arbitrary variables.

Therefore, the connection of PDF and CDF are determined as $CDP = \int_0^1 P(x)dx$. PV implements the sampling design variable x_i while creating an arbitrary number from the range of zero and one. This relates to achieving this variable by calculating the inverse functions of χ_j fixed to inverse (CDF). For determining the optimum individuals under the procedure of compact technique, a comparative of 2 variables is carried out. The 2 variable agents of bats were 2 sampled those carried out in PV. The “winner” refers to the vector with fitness score which is superior to another member and “loser” is demonstrated based on the separate lesser fitness estimation. The 2 return variables, winner as well as loser were main function estimations that relate a novel candidate with preceding global optimum. In order to upgrade, μ , and δ were assumed that dependent upon the subsequent rules. When the mean value of μ was considered as 1, the upgrade rule to all their elements are μ_i^t, δ_i^t set forwarded to $\mu_i^{t+1}, \delta_i^{t+1}$ as provided from the formulated under:

$$\mu_i^{t+1} = \mu_i^t + \frac{1}{N_p} (\text{winner}_i - \text{loser}_i) \quad (6)$$

where N_p refers to the virtual populations. With respect to δ value, the upgrade rule of all elements is provided as follows:

$$\delta_i^{t+1} = \sqrt{(\delta_i^t)^2 + (\mu_i^t)^2 - (\mu_i^{t+1})^2} + \frac{1}{N_p} (\text{winner}_i^2 - \text{loser}_i^2) \quad (7)$$

$$[\text{winner}, \text{loser}] = \text{compete}(x_{\text{best}}, x^{t+1}) \quad (8)$$

Generally, the probabilistic method to CBA was utilized for representing the bat solutions set whereas if the position or velocity were storing, but a recently created candidate was saved. The CBA approach resolves an FF for attaining enhanced classification performance. It defines a positive integer to represent the better performance of the candidate solutions. In this study, the minimization of classification error rate is considered as the fitness function, as given in Eq. (9). The optimal solution has a minimal error rate and the worse solution attains an increased error rate.

$$\text{fitness}(x_i) = \text{Classifier Error Rate}(x_i) = \frac{\text{number of misclassified objects}}{\text{Total number of objects}} * 100 \quad (9)$$

4 Experimental Validation

This section inspects the object detection and classification results of the PFPN-ADT model on distinct datasets such as USCD anomaly detection dataset [22] and other datasets [23,24] (includes City scapes, ADE20K, COCO). A few sample Pedestrian images are demonstrated in Fig. 2.

Fig. 3 shows the visualization result analysis of the PFPN-ADT model on test images. The figure reported that the PFPN-ADT model has effectually identified the objects that exist in the frame. The images in the first row denote the original images and the respective segmented images are offered in the second row. Tab. 1 demonstrates the Area Under Curve (AUC) and Equal Error Rate (EER) investigation of the PFPN-ADT model on the test pedestrian-1 and pedestrian-2 datasets. Fig. 4 indicates the comparative AUC examination of the PFPN-ADT model with existing models on pedestrian-1 and pedestrian-2 datasets. The results represented that the PFPN-ADT model has resulted in increased values of AUC on both datasets. For instance, with pedestrian-1 dataset, the PFPN-ADT model has offered higher AUC of 92.35% whereas the Adam, RBM, ST-AE, ConvAE, ST-CaAE, and EDL-VAD models have obtained lower AUC of 64.43%, 68.99%, 90.48%, 79.29%, 90.55% and 84.94% respectively. In addition, with pedestrian-2 dataset, the PFPN-ADT method has presented high AUC of 94.23% while the

Adam, RBM, ST-AE, ConvAE, ST-CaAE and EDL-VAD systems have gained lesser AUC of 62.42%, 84.51%, 87.59%, 90.19%, 93.23% and 91.88% correspondingly.

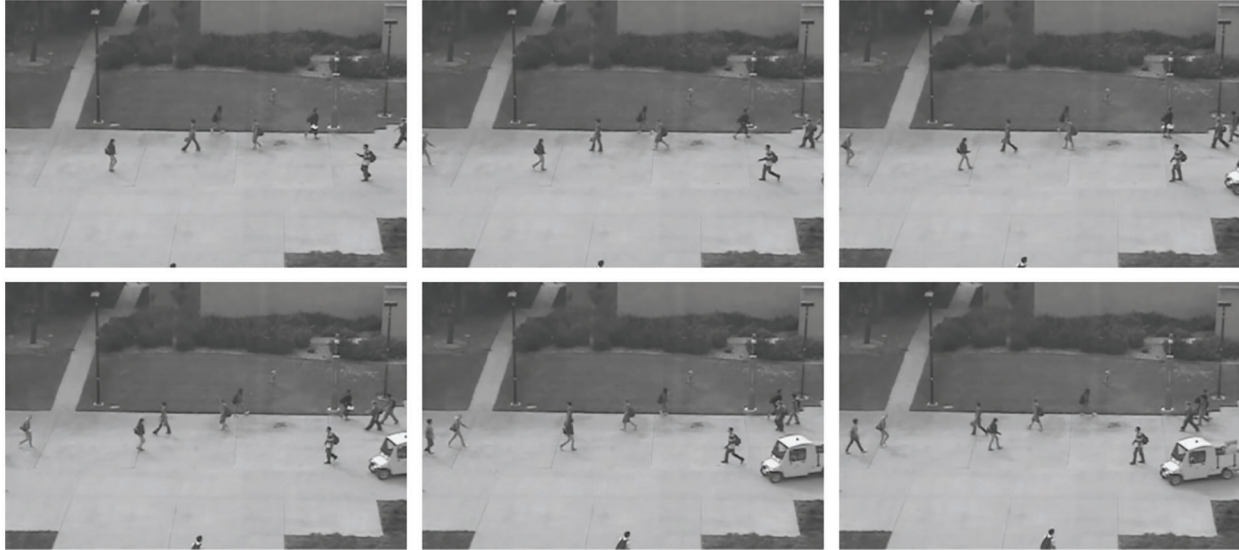


Figure 2: Sample pedestrian images

Fig. 5 shows the comparative EER inspection of the PFPN-ADT model with existing models on pedestrian-1 and pedestrian-2 datasets. The results signified that the PFPN-ADT model has accomplished least values of AUC on both datasets. For instance, with pedestrian-1 dataset, the PFPN-ADT model has gained lower AUC of 9.68% whereas the Adam, RBM, ST-AE, ConvAE, ST-CaAE and EDL-VAD models have obtained higher AUC of 36.14%, 34.29%, 10.65%, 27.93%, 18.41% and 22.48% respectively. Also, with pedestrian-2 dataset, the PFPN-ADT approach has attained lesser AUC of 10.16% while the Adam, RBM, ST-AE, ConvAE, ST-CaAE and EDL-VAD systems have attained maximum AUC of 41.88%, 15.16%, 10.01%, 21.65%, 12.67% and 14.01% correspondingly.

Tab. 2 and Fig. 6 inspects the Running Time (RT) examination of the PFPN-ADT model on pedestrian-1 and pedestrian-2 dataset. The results showcased that the PFPN-ADT model has led to minimal RT on both datasets. For instance, on pedestrian-1 dataset, the PFPN-ADT model has attained lower RT of 0.95 s whereas the MDT, AMDN, ST-CNN, AED and ICN models have obtained higher RT of 24.85, 5.46, 3.37, 1.73 and 1.21 s respectively. Moreover, on pedestrian-2 dataset, the PFPN-ADT model has reached minimal RT of 0.92 s whereas the MDT, AMDN, ST-CNN, AED and ICN models have accomplished maximum RT of 27.58, 7.91, 3.59, 1.56 and 1.18 s respectively.

Tab. 3 provides a detailed anomaly detection and classification outcome of different methods on three benchmark datasets. Fig. 7 depicts a comparative result analysis of the PFPN-ADT model with existing methods on Cityscapes dataset. The results indicated that the MaskRCNN model has shown ineffectual outcomes with the Average Precision (AP), Panoptic Quality (PQ), Segmentation Quality (SQ), and Recognition Quality (RQ) of 30.21%, 51.61%, 77.11% and 61.94% respectively. Followed by, the Mask RCNN-COCO technique has resulted in slightly improved AP, PQ, SQ and RQ values of 35.25%, 53.95%, 77.52% and 66.42% respectively. However, the PFPN-ADT model has resulted in higher AP, PQ, SQ, and RQ of 39.67%, 57.65%, 82.54% and 76.32% respectively.

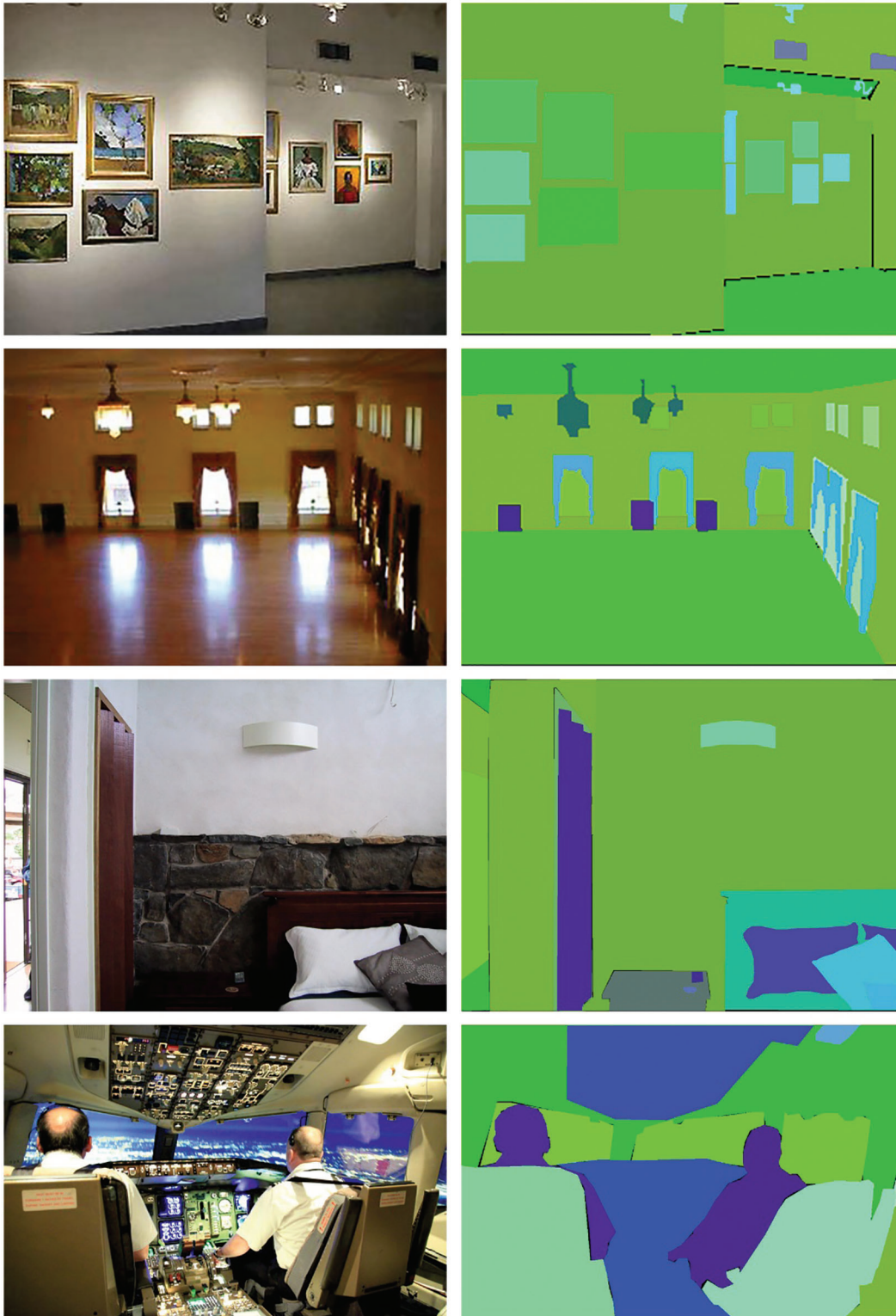


Figure 3: First row original images and second row segmented images

Table 1: Result analysis of PFPN-ADT technique in terms of AUC and EER under two pedestrian datasets

Methods	Pedestrian-1		Pedestrian-2	
	AUC (%)	EER (%)	AUC (%)	EER (%)
Adam model	64.43	36.14	62.42	41.88
RBM model	68.99	34.29	84.51	15.16
ST-AE model	90.48	10.65	87.59	10.01
ConvAE	79.29	27.93	90.19	21.65
ST-CaAE	90.55	18.41	93.23	12.67
EDL-VAD	84.94	22.48	91.88	14.01
PFPN-ADT	92.35	09.68	94.23	10.16

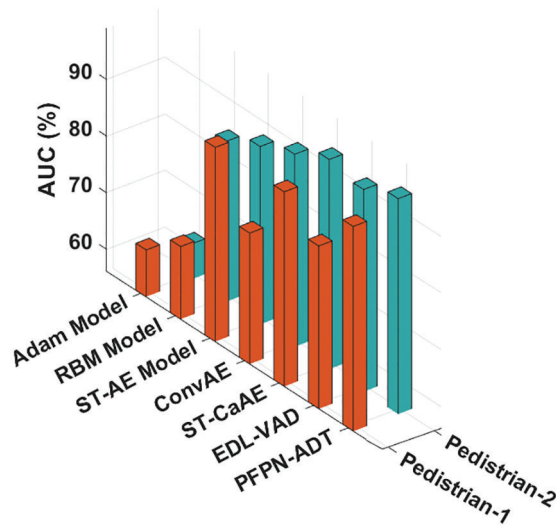


Figure 4: AUC analysis of PFPN-ADT technique under two pedestrian datasets

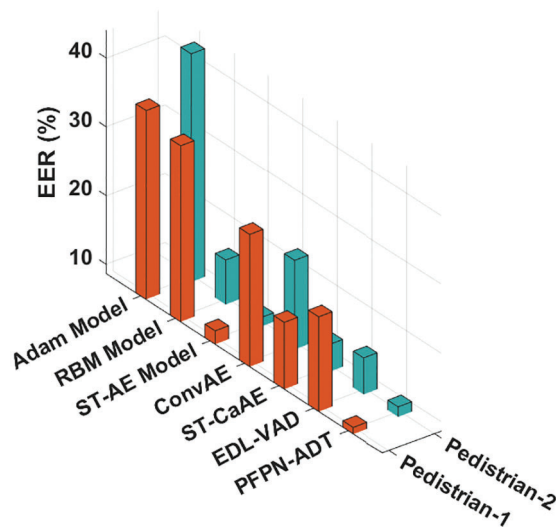
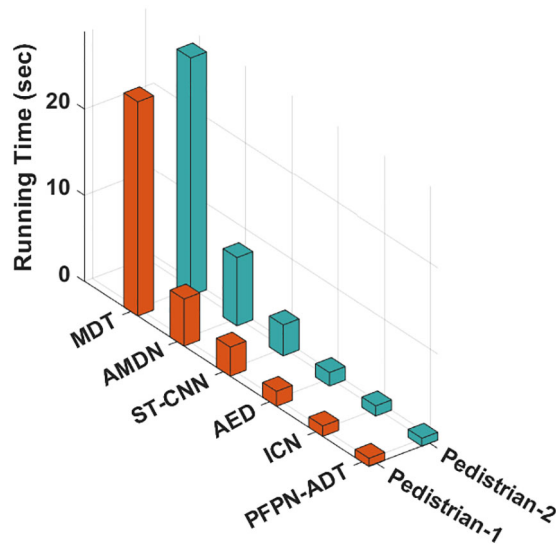


Figure 5: EER analysis of PFPN-ADT technique under two pedestrian datasets

Table 2: Running time analysis of PFPN-ADT technique with recent approaches

Methods	Running time (s)	
	Pedestrian-1	Pedestrian-2
MDT	24.85	27.58
AMDN	05.46	07.91
ST-CNN	03.37	03.59
AED	01.73	01.56
ICN	01.21	01.18
PFPN-ADT	01.19	01.14

**Figure 6:** Running time analysis of PFPN-ADT technique with recent approaches**Table 3:** Comparative analysis of PFPN-ADT technique three benchmark datasets

Methods	AP	PQ	SQ	RQ
Cityscapes				
PFPN-ADT	39.67	57.65	82.54	76.32
Mask RCNN-COCO	36.25	53.95	77.52	66.42
Mask RCNN	30.21	51.61	77.11	61.94
ADE20k				
PFPN-ADT	35.16	51.88	85.23	53.62
Mask RCNN-COCO	28.74	40.43	79.66	48.12
Mask RCNN	24.80	35.86	78.17	41.96
Vistas				
PFPN-ADT	38.27	56.48	67.11	70.68
Mask RCNN-COCO	33.09	52.88	63	67.18
Mask RCNN	30.44	50.71	61.6	65.11

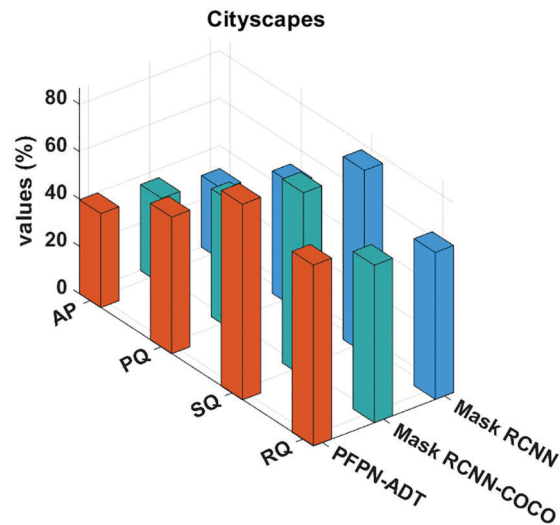


Figure 7: Comparative analysis of PFPN-ADT technique under cityscapes dataset

Fig. 8 shows a comparative analysis of the PFPN-ADT technique with current approaches on ADE20k dataset. The result indicates that the MaskRCNN system has revealed unsuccessful outcomes with the AP, PQ, SQ and RQ of 24.80%, 35.86%, 78.17% and 41.96% correspondingly. Next, the Mask RCNN-COCO method has resulted in slightly enhanced AP, PQ, SQ and RQ values of 28.74%, 40.43%, 79.66% and 48.12% correspondingly. But, the PFPN-ADT approach has resulted in maximal AP, PQ, SQ and RQ of 35.16%, 51.88%, 85.23% and 53.62% correspondingly.

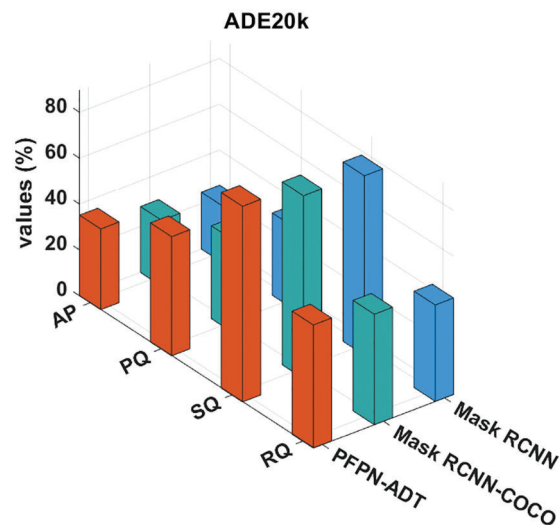


Figure 8: Comparative analysis of PFPN-ADT technique under ADE20k dataset

Fig. 9 shows a comparative analysis of the PFPN-ADT approach with current systems on Vistas dataset. The outcomes showed that the MaskRCNN method has revealed unsuccessful outcomes with the AP, PQ, SQ, and RQ of 30.44%, 50.71%, 61.6% and 65.11% correspondingly. After that, the Mask RCNN-COCO system has resulted in slightly enhanced AP, PQ, SQ and RQ values of 33.09%, 52.88%, 63%

and 67.18% correspondingly. But, the PFPN-ADT system has resulted in maximum AP, PQ, SQ and RQ of 38.27%, 56.48%, 67.11% and 70.68% correspondingly.

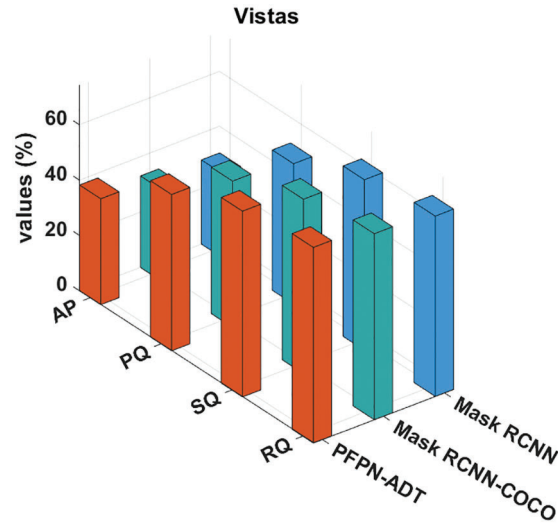


Figure 9: Comparative analysis of PFPN-ADT technique under Vistas dataset

Finally, a detailed CDF examination of the PFPN-ADT model takes place on three datasets namely Cityscapes, ADE20k and Vistas under distinct Intersection Over Union (IOU) as shown in [Tab. 4](#) and [Fig. 10](#) [25–27]. The results depicted that the CDF values tend to increase with a rise in IOU. For instance, with 20 IOU, the PFPN-ADT model has offered CDF of 1.13%, 2.45% and 4.42% on Cityscapes, ADE20k and Vistas datasets respectively. Moreover, with 60 IOU, the PFPN-ADT approach has presented CDF of 10.34%, 14.95% and 25.49% on Cityscapes, ADE20k and Vistas datasets correspondingly.

Table 4: CDF analysis of PFPN-ADT model under different IOU

IOU	Cumulative Distribution Function (%)		
	Cityscapes	ADE20k	Vistas
0	1.13	1.13	3.10
20	1.13	2.45	4.42
40	3.10	5.08	8.37
60	10.34	14.95	25.49
80	41.29	48.53	59.72
100	98.56	98.56	99.87

Furthermore, with 100 IOU, the PFPN-ADT system has presented CDF of 98.56%, 98.56% and 99.87% on Cityscapes, ADE20k and Vistas datasets correspondingly. Therefore, it is exhibited that the PFPN-ADT model has the capability of attaining maximum object detection and classification performance.

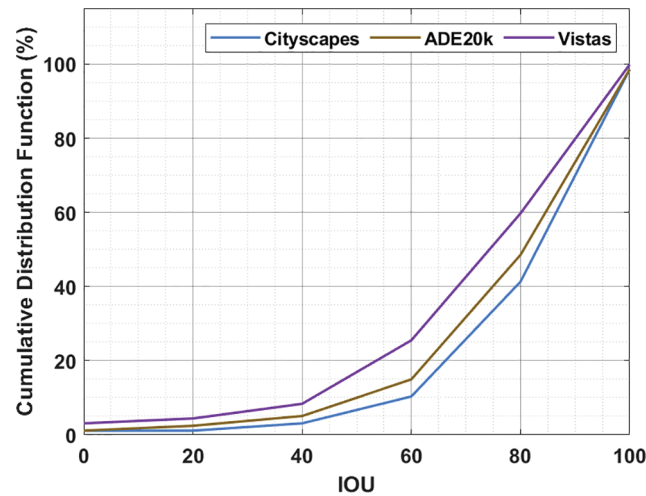


Figure 10: CDF analysis of PFPN-ADT model under different IOU

5 Conclusion

In this study, a new PFPN-ADT system has been developed for the recognition and classification of different anomalies present in the pedestrian walkway like vehicles and skaters, etc. The PFPN-ADT model encompasses PFPN based object detection, SAE based object classification, and CBA based parameter optimization. The design of PFPN and CBA assists in accomplishing enhanced object detection and classification performance. For ensuring the enhanced results better anomaly detection performance of the PFPN-ADT technique, a comparison study is made using UCSD Anomaly data and other benchmark datasets (such as Cityscapes, ADE20K and COCO), and the outcomes are compared with the Mask RCNN and Faster CNN models. The simulation outcome demonstrated the enhanced performance of the PFPN-ADT technique over the other methods.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. K. Pal, A. Pramanik, J. Maiti and P. Mitra, "Deep learning in multi-object detection and tracking: State of the art," *Applied Intelligence*, vol. 51, no. 9, pp. 6400–6429, 2021.
- [2] P. S. Prakash, R. R. Rajakshmi and T. Kumaravel, "Object detection in surveillance video," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 9, pp. 477–487, 2021.
- [3] S. Burde and S. V. Budihal, "Multiple object detection and tracking using deep learning," in *Proc. Int. Conf. on Communication, Circuits and Systems*, Singapore, Springer, pp. 257–263, 2021.
- [4] S. Jha, C. Seo, E. Yang and G. P. Joshi, "Real time object detection and tracking system for video surveillance system," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3981–3996, 2021.
- [5] B. Xiao and S. C. Kang, "Vision-based method integrating deep learning detection for tracking multiple construction machines," *Journal of Computing in Civil Engineering*, vol. 35, no. 2, pp. 4020071, 2021.
- [6] K. Kalirajan, V. Seethalakshmi, D. Venugopal and K. Balaji, Deep learning for moving object detection and tracking. In: *Examining the Impact of Deep Learning and IoT on Multi-Industry Applications*. United States: IGI Global, pp. 136–163, 2021.

- [7] D. R. Niranjana and B. C. VinayKarthik, "Deep learning based object detection model for autonomous driving research using CARLA simulator," in *Proc. 2nd Int. Conf. on Smart Electronics and Communication (ICOSEC)*, Trichy, India, IEEE, pp. 1251–1258, 2021.
- [8] D. H. Lee, "CNN-based single object detection and tracking in videos and its application to drone detection," *Multimedia Tools and Applications*, vol. 80, no. 26, pp. 34237–34248, 2021.
- [9] B. Xiao and S. C. Kang, "Development of an image data set of construction machines for deep learning object detection," *Journal of Computing in Civil Engineering*, vol. 35, no. 2, pp. 5020005, 2021.
- [10] L. Jiao, R. Zhang, F. Liu, S. Yang, B. Hou *et al.*, "New generation deep learning for video object detection: A Survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3195–3215, 2022.
- [11] D. H. Ye, J. Li, Q. Chen, J. Wachs and C. Bouman, "Deep learning for moving object detection and tracking from a single camera in unmanned aerial vehicles (UAVs)," *Electronic Imaging*, vol. 2018, no. 10, pp. 466–478, 2018.
- [12] Y. W. Hsu, T. Y. Wang and J. W. Perng, "Passenger flow counting in buses based on deep learning using surveillance video," *Optik*, vol. 202, pp. 163675, 2020.
- [13] H. Wang, P. Wang and X. Qian, "MPNET: An end-to-end deep neural network for object detection in surveillance video," *IEEE Access*, vol. 6, pp. 30296–30308, 2018.
- [14] S. Mane and S. Mangale, "Moving object detection and tracking using convolutional neural networks," in *Proc. Second Int. Conf. on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, IEEE, pp. 1809–1813, 2018.
- [15] E. Avcı and Y. O. Avcı, "Moving vehicle detection and tracking at roundabouts using deep learning with trajectory union," *Multimedia Tools and Applications*, vol. 81, no. 5, pp. 6653–6680, 2022.
- [16] K. S. Yadav, K. Anish Monsley, R. H. Laskar, S. Misra, M. K. Bhuyan *et al.*, "A selective region-based detection and tracking approach towards the recognition of dynamic bare hand gesture using deep neural network," *Multimedia Systems*, vol. 82, no. 3, pp. 1–19, 2022.
- [17] A. Kirillov, R. Girshick, K. He and P. Dollar, "Panoptic feature pyramid networks," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 6399–6408, 2017.
- [18] A. Kirillov, K. He, R. Girshick, C. Rother and P. Dollar, "Panoptic segmentation," in *Proc. Conf. on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp. 9404–9413, 2019.
- [19] V. Navya and K. A. Patil, "Identification of anomalies in images using CNN and autoencoders techniques," in *Proc. Int. Conf. on Recent Trends in Computing*, Singapore, Springer, pp. 307–316, 2022.
- [20] N. Krishnaraj and S. Sangeetha, "A study of data privacy in internet of things using privacy preserving techniques with its management," *International Journal of Engineering Trends and Technology*, vol. 70, no. 2, pp. 43–52, 2022.
- [21] J. S. Pan and T. K. Dao, "A compact bat algorithm for unequal clustering in wireless sensor networks," *Applied Sciences*, vol. 9, no. 10, pp. 1973, 2019.
- [22] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, pp. 1–12, 2010.
- [23] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso *et al.*, "Scene parsing through ade20k dataset," in *Proc. Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 633–641, 2017.
- [24] N. Krishnaraj, M. Elhoseny, M. Thenmozhi, M. M. Selim and K. Shankar, "Deep learning model for real-time image compression in Internet of Underwater Things (IoUT)," *Journal of Real-Time Image Processing*, vol. 17, no. 6, pp. 2097–2111, 2020.
- [25] C. Wu, S. Shao, C. Tunc, P. Satam and S. Hariri, "An explainable and efficient deep learning framework for video anomaly detection," *Cluster Computing*, vol. 12, no. 1, pp. 1–23, 2021.
- [26] K. Yang, J. Jiang and Z. Pan, "Mixed noise removal by residual learning of deep cnn," *Journal of New Media*, vol. 2, no. 1, pp. 1–10, 2020.
- [27] X. R. Zhang, J. Zhou, W. Sun and S. K. Jha, "A lightweight CNN based on transfer learning for COVID-19 diagnosis," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1123–1137, 2022.