Safety Science 142 (2021) 105356

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

An automated deep learning based anomaly detection in pedestrian walkways for vulnerable road users safety

Irina V. Pustokhina [a], Denis A. Pustokhin [b], Thavavel Vaiyapuri [c], Deepak Gupta [d], Sachin Kumar [e], K. Shankar [f,*]

[a] Department of Entrepreneurship and Logistics, Plekhanov Russian University of Economics, 117997 Moscow, Russia
[b] Department of Logistics, State University of Management, 109542 Moscow, Russia
[c] College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Saudi Arabia
[d] Department of Computer Science & Engineering, Maharaja Agrasen Institute of Technology, Delhi, India
[e] Department of Computer Science, South Ural State University, Chelyabinsk, Russian Federation
[f] Department of Computer Applications, Alagappa University, Karaikudi, India

ARTICLE INFO

Keywords:
Anomaly detection
Pedestrian walkways
Deep learning
Safety
Mask RCNN

ABSTRACT

Anomaly detection in pedestrian walkways is an important research topic, commonly used to improve the safety of pedestrians. Due to the wide utilization of video surveillance systems and the increased quantity of captured videos, the traditional manual examination of labeling abnormal events is a tiresome task. So, an automated surveillance system that detects anomalies becomes essential among computer vision researchers. Presently, the development of deep learning (DL) models has gained significant interest in different computer vision processes namely object classification and object detection, and these applications were depending on supervised learning that required labels. Therefore, this paper develops an automated deep learning based anomaly detection technique in pedestrian walkways (DLADT-PW) for vulnerable road user's safety. The goal of the DLADT-PW model is to detect and classify the various anomalies that exist in the pedestrian walkways such as cars, skating, jeep, etc. The DLADT-PW model involves preprocessing as the primary step, which is applied for removing the noise and raise the quality of the image. In addition, mask region convolutional neural network (Mask-RCNN) with densely connected networks (DenseNet) model is employed for the detection process. To ensure the better anomaly detection performance of the DLADT-PW technique, an extensive set of simulations were performed and the outcomes are investigated under distinct aspects. The obtained experimental values confirmed the superior characteristics of the DLADT-PW technique by achieving a maximum detection accuracy.

1. Introduction

Annually, more than 270 000 pedestrians lose their lives on the world's roads. The capacity to respond to pedestrian safety is an important component of efforts to prevent road traffic injuries. Pedestrian collisions, like other road traffic crashes, should not be accepted as inevitable because they are, in fact, both predictable and preventable. Recent technological advances like computer vision (CV), surveillance cameras (CCTV), etc. can be used to protect pedestrians and promote safe walking require an understanding of the nature of risk factors for pedestrian crashes. This study aims to ensure the safety of pedestrians using computer vision techniques. An extensive application of surveillance cameras (CCTV) in public places led the CV centric model to learn the reputation over the CV research team. The captured visual data is composed of enriched details which are accurate when compared with alternate data sources like GPS, mobile communication, radar signals, and so on. Also, it plays a major role in forecasting congestion, accidents, and some other abnormal activities by gathering details regarding the condition of road traffic. The use of CCTV finds helpful in several real time applications (Zhang et al., 2018; Cocca et al., 2016; Wester and Giesecke, 2019; Tsai, 2014; Rahouti et al., 2020) like grade-crossing-trespassing, industrial safety management, accidental fall, etc.

Numerous computer vision depends on works have been proposed by concentrating on operations like data acquisition, feature extraction,

* Corresponding author.
E-mail addresses: ivpustokhina@yandex.ru (I.V. Pustokhina), dpustokhin@yandex.ru (D.A. Pustokhin), t.thangam@psau.edu.sa (T. Vaiyapuri), deepakgupta@mait.ac.in (D. Gupta), sachinagnihotri16@gmail.com (S. Kumar), drkshankar@ieee.org (K. Shankar).

https://doi.org/10.1016/j.ssci.2021.105356
Received 22 January 2021; Received in revised form 2 May 2021; Accepted 24 May 2021
Available online 10 June 2021

scene learning, activity learning, behavioral learning, and so forth. The basic aim of these studies is to compute the operations like scene detection, video processing models, anomaly prediction approaches, vehicle prediction and observation, multi camera-relied schemes and challenges, activity examination, traffic observation, human behavior learning, and so on. Here, anomalous prediction is considered to be a sub-domain of behavior learning from the captured visual scenes. The accessibility of video from public places has resulted in the simulation of video analysis as well as anomalous prediction (Rahouti et al., 2020). Moreover, anomalous prediction approaches understand the common behavior by the training process. Any significant change from normal behavior is considered to be anomalous. The existence of vehicles on pathways, unexpected dispersion of people from a crowd, person faints whereas walking, jaywalking, signal bypassing at a traffic junction, U-turn of vehicles in red signals are some of the common examples of anomalies.

In general, anomaly prediction approaches apply unsupervised as well as semi-supervised learning. An important aim of this work is for finding the anomaly prediction schemes applied in road traffic cases and concentrates on the utilities like vehicles, trespassers, atmosphere, and communication. It has been pointed that, the scope of this has to enclose the nature of input data as well as the representations, possibility of supervised learning, class of abnormalities, the capability of the systems in application content, anomaly prediction results as well as termination criteria. The anomaly prediction mechanism is operated by understanding the common data patterns for developing a public profile. When the general patterns are defined, anomalies could be predicted using newly developed schemes. Hence, the simulation of the model is a label that predicts whether data is abnormal or healthy.

Recently, diverse models were deployed for computing pedestrian prediction which suits the bounding boxes for a pedestrian available in an image. It has gained maximum attention from the developers of computer vision and the significant element for diverse human-based domains such as driverless cars, automated traffic signaling, person examination, etc. However, the predefined models are unfit for resolving the complexity of a model named scaling problem that remains the same and causes the outcome of pedestrian detection approach. The traditional approaches have managed to solve the scaling problem on the 2D scale. First, brute-force data is augmented to improve the capability of the scale-invariance model. Followed by, a single method with multiple scale filters was employed in all samples with diverse sizes. However, the presence of intra-class variance of maximum and tiny samples is complicated to overcome the significantly varied feature responses along with individual approaches. To make use of drastically differing attributes with varied scales, the divide-and-conquer paradigm can be applied (Gong et al., 2014) for resolving the complicated scale variance problem.

Ultimately, Deep Learning (DL) relied on anomaly prediction methods are deployed. Initially, Convolution Neural Network (CNN) has been employed and categorized the presence of objects. It has experienced few issues like massive spatial locations as well as aspect ratios of objects from an image. In order to overcome these problems, a number of regions have to be selected and results in processing complexity. Thus, region-based CNN (R-CNN) and YOLO have been established to find the incidence at a robust rate. Here, a novel approach has been developed and overcome the problems involved in selecting a maximum number of regions and employed a selective search mechanism to extract images called region proposals. Finally, the selective search model has generated a maximum number of regions.

In order to enhance the safety of pedestrians, this paper designs a novel DL based anomaly detection technique in pedestrian walkways (DLADT-PW). The DLADT-PW model aims to recognize and categorize the dissimilar anomalies present in the pedestrian walkways such as cars, skating, jeep, etc. The DLADT-PW model includes preprocessing as the primary step, which is applied to eradicate the noise and increase the quality of the image. Besides, mask region convolutional neural network (Mask-RCNN) with densely connected networks (DenseNet) model is applied for the detection process. For verifying the superior anomaly detection performance of the DLADT-PW technique, a wide set of simulations were accomplished and the results are inspected under distinct aspects.

The rest of the paper is organized as follows. Section 2 briefs the existing works and section 3 discusses the proposed model. Then, section 4 validates the performance of the proposed model and section 5 concludes the paper.

2. Literature review

This section intends to survey an extensive set of available hand-crafted feature based anomaly detection techniques and deep learning based anomaly detection techniques.

2.1. Hand-Crafted features based method

In general, 3 components could be filtered from hand-engineered features relied on the anomalous prediction approach. In case of a feature extraction system, diverse feature descriptions have been developed (Yang et al., 2020). At this point, low-level trajectory attributes from series of images have been applied to define the normal movement patterns. But, the above-mentioned approaches are concentrated on anomaly affected by a crowd rather than a single object is considered as a basic element. Hence, the trajectory features are depending upon crowd monitoring and these approaches are unfit in handling single object anomaly prediction (Alqaralleh et al., 2020). Additionally, the trajectory features have minimum-level spatio-temporal features like the histogram of oriented flows (HOF) as well as the histogram of oriented gradients (HOG). Kratz and Nishino (Kratz and Nishino, 2009) utilized the dispersion of spatiotemporal gradients to demonstrate the appropriate motion details in local spatiotemporal motion. In (Xu et al., 2014), the motion feature depicted by the histogram of optical flow is employed as a low-level feature for motion-pattern definition.

Kim and Grauman (Kim and Grauman, 2009) utilized the mixture of probabilistic principal component analyzers (MPPCA) methods for defining the local activity patterns using optical flow as low-level metrics. Mahadevan et al. (Mahadevan et al., 2010) examined a technology for normal crowd features that relied on mixtures of dynamic textures (MDT) and Li et al. (Li et al., 2014) employed a Conditional Random Field (CRF) to combine the results according to the given application. For modeling, the existence, as well as motion features from PCA, Feng et al. (Feng et al., 2017), established a deep Gaussian mixture model (GMM). Moreover, few sparse coding approaches were employed for encoding the normal patterns. Next, the normal dictionary has been learned from over complete normal basis set and the sparse reforming cost has been applied for measuring the common feature of the testing sample. Eventually, the training and testing process can be triggered by, Lu et al. (2013) with the help of several dictionaries for encoding the normal size-invariant blocks from multiscale frames. Yu et al. (2017) visualized the low-rank feature of bases from dictionary learning state, afterward, a weighted sparse reformation scheme has been applied for measuring the abnormality of samples.

2.2. Deep learning based method

Recently, DL frameworks are employed in massive computer vision process effectively, and in anomaly, prediction works (Alqaralleh et al., 2020). Mostly, convolutional AE or fully convolutional systems have been applied for reforming a novel group of frames. In case of sequence video frames with no abnormalities, Liu et al. (2018) have trained Fully Convolutional Network (FCN) approach which mimics the U-Net for predicting the consecutive frame. Followed by, the deviations among predicted frame and corresponding ground truth frames were applied

for predicting the anomalies in the detection state. Ribeiro et al. (2018) utilized the outcome of convolutional AE which has been assumed for redeveloping input frame sequences. Since the AE is trained under the application of normal video sequences, reconstruction error has been employed as an anomaly value. Therefore, the better applicability, as well as normalization of Deep Neural Network (DNN), the consideration of anomalous events, may accelerate maximum reconstruction errors. Hence, the main objective of these models is extracting features using AE and predicting the anomalies by probability estimation of features.

Sabokrou et al. (Sabokrou et al., 2017) developed a deep convolutional neural network (DCNN) along with the kernels equipped by using sparse AE (SAE). Considering the cubic patches obtained from actual images as inputs, feature maps from 3 middle and final layers have been induced as the Gaussian classification model. In Xu et al. (Xu et al., 2017), 3 stacked denoising AE have been projected for learning the spatial features, temporal features, and the mixture of these 2 models. Next, 3 one-class SVM methods are employed for estimating the learned features and examine the anomaly values. In addition, Sabokrou et al. (Sabokrou et al., 2018) have projected a pre-trained CNN and intercepted it as a feature extractor into FCN which is capable of extracting the features for receptive field with no cropping the input frames as patches.

3. The proposed DLADT-PW technique

The workflow involved in the presented DLADT-PW technique is given here. As depicted, the surveillance video is primarily converted into a set of frames, and anomalies are detected in each frame. Next, the preprocessing is performed to improve the quality of the image. Followed by, the Mask RCNN model is applied for the detection of anomalies and DenseNet 169 model is utilized as the baseline network for Mask RCNN. At last, the anomalies that exist in the frame are successfully identified and classified.

3.1. Preprocessing

In general, the collected data is complicated and composed of inappropriate images, blurred images, and noisy images. Followed by, the data is subjected to clean, smooth, and label so that the dataset quality is enhanced. At this point, eliminate the noisy images. Next, Median Filtering is used for smoothening the image noise caused by several unwanted objects on target objects. An image histogram is one of the typical approaches used for data cleaning. The actual purpose of this model for transforming the image into a histogram and apply a correlation coefficient model for identifying the image homogeneity. The estimation formula of the correlation coefficient is given below:

$$r(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}[x]\mathrm{Var}[y]}}, \tag{1}$$

where x, y indicates the histogram outcome of 2 images, $\mathrm{Var}[x]$ defines the covariance of x, $\mathrm{Var}[y]$ refers the covariance of y, and $\mathrm{Cov}(x,y)$ represents the covariance from x and y. Besides, the score of the applied function is ranked from [-1 to 1]. Moreover, the estimated outcome has most of the similarities between the 2 images. Here, the histogram model has been applied for removing the gathered images. In prior to applying the histogram, eliminate the irregular images which are often irrelevant. Afterward, select an appropriate image for a class and estimate the correlation coefficient among correct image and residual images.

A major variance between image and noise is considered to be the extension of the gray level. Therefore, the visual obstacle of an image is caused by a drastic difference between the gray level of noise and the corresponding gray level (Li et al., 2020). Thus, an image smoothing mechanism is applied for removing the noise with the help of grayscale variations.

3.2. Mask R-CNN

In general, Mask RCNN is an elegant, flexible, and common approach used for object prediction, and instance segmentation which is capable of predicting the objects available in an image at the time of generating a high-dimension segmentation mask. Feature Pyramid Networks (FPNs) are employed for object prediction and the first block architecture of Mask R-CNN is applied for feature extraction. Hence, Regional Proposal Network (RPN), is considered to be the second block of Mask RCNN which distributes the convolutional features in conjunction with the prediction system and enables the cost-free RP. The RPN is also employed to Mask RCNN rather than using selective search and the RPN distribution of convolution feature in full map with the detection system. It is also capable of predicting boundary location and object values in every position and FCN.

In order to enhance the forecast accuracy of the method, Mask RCNN applies the bilinear interpolation approach named region of interest (ROI) according to the Faster RCNN. Also, ROI aligns layer eliminates the harsh quantization of the ROI pool as well as aligns the obtained features properly with the input (Xu et al., 2020). Afterward, this approach of ROI alignment is applied for computing the accurate measures of input features based on bilinear interpolation at regularly sampled positions in all ROI bins to accumulate the simulation outcome. Mask R-CNN is suitable in computing 3 processes like target detection, prediction, and segmentation. Here, when the image is conveyed by FPN, 5 sets of feature maps are produced with different sizes, and the candidate frame region is emanated by the RPN. Classification detection in Mask RCNN is relevant with mask branch and applied for gaining the spatial structure of object with the help of pixel-to-pixel organization from convolutional layers which undergoes encoding. By means of potential misalignment among input as well as feature maps without ROI Pooling, Roi Align has been applied in the Mask RCNN by applying bilinear interpolation to enhance the model accuracy.

3.2.1. RPn

In feature maps from convolutional layers and network proceeds convolutional process on 3*3 pixels sliding window. A point in feature maps emanates feature codes for respective window regions that concern the minimum-dimensional feature codes of dimensions from Mask RCNN. Followed by, ranking of classification values from initial regression feature boxes which are decided, and the values of relevant coordinates undergo decoding as accurate coordinates by the given Eqs. (2) and (3):

$$t_x = (x - x_a)/w_a, t_y = (-y_a)/h_a \tag{2}$$

$$t_w = \log(w/w_a), t_h = (h/h_a) \tag{3}$$

Where (x_a, y_a) implies the manages of the center of anchor and (w_a, h_a) denotes the height as well as the width of the anchor. (x, y) depicts the direct of middle forecasted ROI in actual image and (w, h) defines the height and width of ROI detected in the ground truth image. (t_x, t_y) signifies the regression score of coordinates and (t_w, t_h) represents the regression score of the height and width on the feature map. In particular, when the measure of intersection-over-union (IoU) from the detected bounding boxes from ROI along with ground truths are maximum than the defined threshold where the targets in ROI are considered as a foreground as well as background.

3.2.2. Loss function

In multi-task loss function has been applied in training Mask RCNN with 3 portions namely, classification loss of bounding box, location regression loss of bounding box as well as loss of mask as depicted by the given function.

$$L = L_{cls} + L_{box} + L_{mask} \tag{4}$$

$$L_{cls} = -\log\left[p_i{}^{*}p_i + \left(1-p_i^{*}\right)\left(1-p_i\right)\right] \quad (5)$$

$$L_{box} = r\left(t_i - t_i^{*}\right) \quad (6)$$

$$L_{mask} = Sigmoid(Cls_k) \quad (7)$$

Where p_i denotes the detected probability for ROI in classification loss L_{cls} and p_i^{*} used to ground truth as 1 when the ROI is assumed as foreground or 0 else. t_i denotes the vector of accurate manages to detected bounding box (Eq. (6)) and t_i^{*} refers to the ground truth from position regression loss in which r means the robust loss function to estimate the regression error (Xu et al., 2020). Every ROI detects the result of K^{*}m^2 dimensions by using mask branch and encoding K binary masks along with a resolution of m*m. The loss of mask L_{mask} is assumed as the Average Binary Cross-entropy Loss to perform the sigmoid function on every pixel from ROI. In class $k(Cls_k)$, the mask loss is depicted in Eq. (7).

3.3. DenseNet 169 model

The baseline of the Mask RCNN contains the DenseNet-169 model. The DenseNet structure is developed from ResNet, which is comprised of a building block where it is unified with the former layer. Here, excess merges are employed to learn residuals-based errors. DenseNet has projected the mixture of outcomes obtained from previous layers despite using the combination. Consider the single image x_0 is passed by CNN. This network is composed of L layers, in which non-linear transformation$H_l(\hat{A}\cdot)$ is implemented, where l refers to the layer indexes. $H_l(\hat{A}\cdot)$ means the composite function like Batch Normalization (BN), Rectified Linear Units (ReLU), Pooling, or Conv. A final result of lth layer is represented by x_l. FFNN connects the outcome of lth layer as input for (l + 1)th layer that intends to generate layer transition: $x_l = H_l(x_{l-1})$. ResNets has a skip-connection that bypasses non-linear conversion under the application of a given identity function:

$$x_l = H_l(x_{l-1}) + x_{l-1} \quad (8)$$

The advantages of ResNets are that the gradient controls are directed from recent layers to existing layers and it is accomplished with the help of identity function. Hence, the unification of identity function and simulation outcome of H_l obstructs the data communication. Moreover, data flow is improvised under the application of multiple connectivity patterns (Huang et al., 2017). Fig. 1 implies the outline of the last DenseNet structure. At last, lth layer has gained the feature-maps of advanced layers, $x_0, \cdots, x_{l-1}$, as input:

$$x_l = H_l([x_0, x_1, \cdots, x_{l-1}]), \quad (9)$$

where$[x_0, x_1, \cdots, x_{l-1}]$ denotes s the mixture of feature-maps generated in layers 0, ⋯, $l-1$. DenseNet is mainly applied to managing numerous connectivity. It has been executed with the help of massive inputs of $H_l(\hat{A}\cdot)$ in Eq. (4) as an individual tensor. The integration used in Eq. (4) is non-feasible if there is a prominent modification in feature map size. Also, down-sampling is employed and classify the network as densely connected blocks. The transition layers used in this study are comprised of BN layers, Conv layer, and average pooling layer.

The function H_l offers k feature maps that apply lth layer with $k_0 + k \times (l-1)$ input feature-maps, where k_0 indicates the channels with the input layer. The drastic difference between DenseNet and former networks is, DenseNet is limited with narrow layers and represented by k = 12. A layer has k feature-maps of the corresponding state in which the growth rate generalizes data into a global state. It is noted that 1 × 1 Conv is assumed as bottleneck layer prior to use 3 × 3 Conv limits of input feature-maps, and enhances the computational efficacy. Most of the time, DenseNet is effective and system with bottleneck layer. Fig. 2 demonstrates the layers in DenseNet-169.

4. Experimental validation

The proposed model is simulated using Python 3.6.5 tool. For validation, UCSD Anomaly Detection Dataset (Murugan et al., 2019) is utilized for the training and testing of the proposed model. In UCSD Anomaly Detection Dataset required a group of images taken from a static camera located at an elevation overlooking pedestrian pathways. A crowd density in the pathway is not static as well as ranged from sparse to over-crowd. In normal cases, the video required only pedestrians while the abnormal performances or anomalies contained the effort of non-pedestrian entities in the walkways. Anomalies occur in the videos like bikers, skaters, vehicles, tiny carts, as well as people walking through pathways or in the grass that surrounds it. Details of the dataset are provided in Table 1. Besides, the parameter setting is given as follows. Batch Size: 64, Optimizer: Adam, Epoch: 100, Learning rate: 0.001, and Activation function: ReLU.

Fig. 3 demonstrates a sample set of images from the UCSD dataset. The image contains the pedestrians with some anomalies.

Fig. 4 visualizes the results offered by the presented DLADT-PW technique on the applied UCSD dataset. Fig. 4a shows the test image involving a set of pedestrians with some anomalies. Fig. 4b shows the detection of two anomalies that exist on the applied input frame. The figure notifies that the DLADT-PW technique has effectively identified the anomalies.

Table 2 has portrayed the detection accuracy of anomalies of the proposed DLADT-PW model on the applied test004 video sequence. The resultant table values denoted the proficient anomaly detection performance of the DLADT-PW model. For instance, anomaly 1 in the frames 078, 091, 092, and 110 are detected with the maximum accuracy of 0.95, 0.96, 0.97, and 0.98 respectively. Besides, the anomalies in the rest of the frames such as 113, 115, 125, 142, 146, 147, 148, 150, 178, 179, and 180 are detected with the maximum identical accuracy of 0.99. Similarly, anomaly 2 in frames 125, 142, and 146 are noticed with high

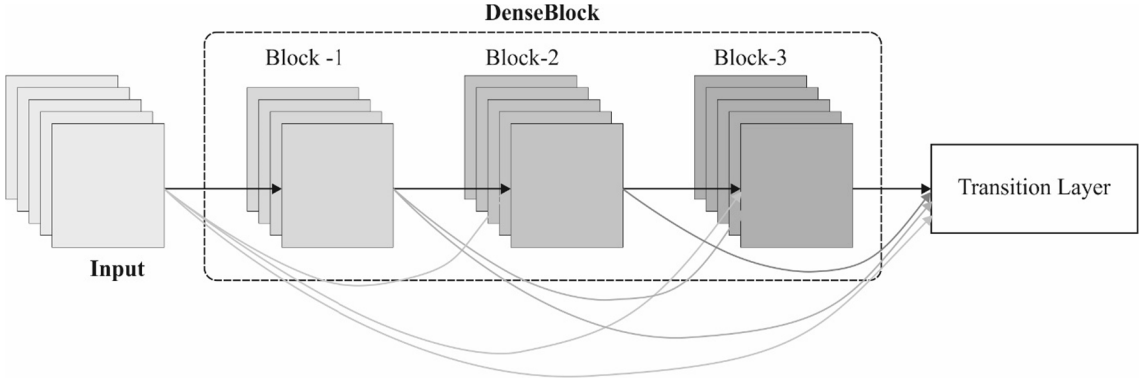

Fig. 1. DenseNet Architecture.

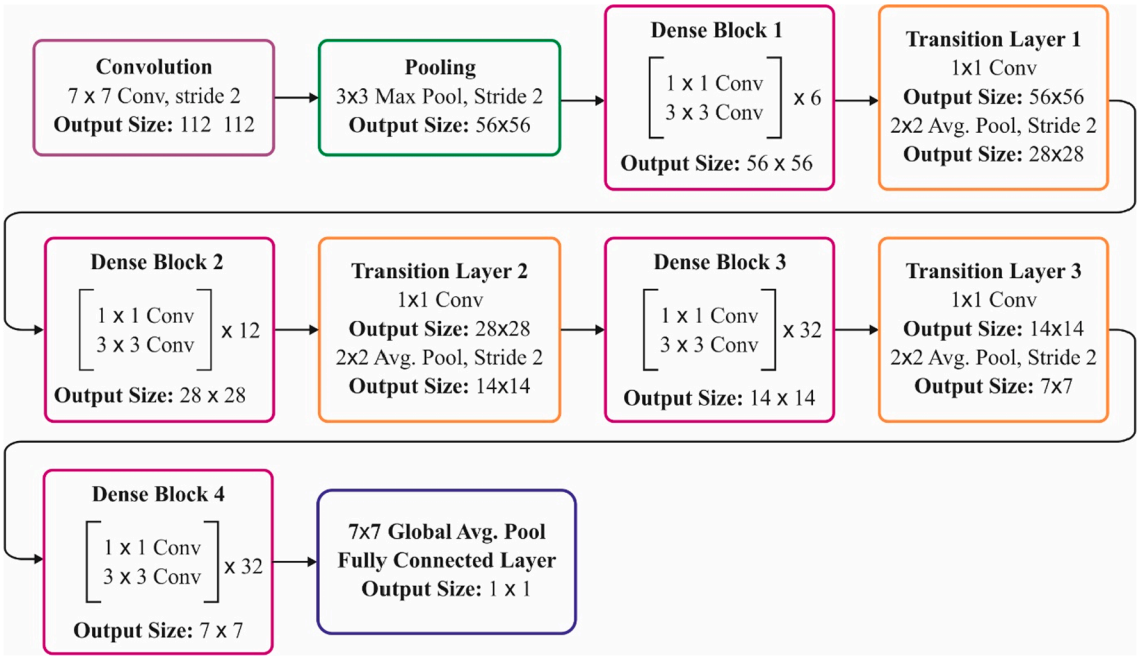

Fig. 2. Layers in DenseNet-169.

Table 1
Description of dataset.

Dataset	Testbed	Frames	Time (sec)
UCSDped2	Test007 Test004	360	12

accuracy of 0.97, 0.98, and 0.97 respectively. Also, the anomalies in the rest of the frames such as 147, 148, 150, 178, 179, and 180 are identified with the maximum identical accuracy of 0.99.

Table 3 and Fig. 5 provided the analysis of the comparative result of the DLADT-PW technique with existing models on the applied Test004 sequence. The values that exist in the table denoted that the SF model has failed to showcase effective detection performance over all the other methods. At the same time, the MDT and MPPCA models have depicted slightly improved outcomes over the SF model. Concurrently, the Fast R-CNN model has demonstrated moderate outcome whereas a near-optimal detection rate is accomplished by the RS-CNN model. However, the presented DLADT-PW model has resulted in maximum detection performance over all the other compared methods. For instance, on the applied frame of 040, the DLADT-PW model has obtained a maximum accuracy of 0.950 whereas the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.940, 0.819, 0.768, 0.744, and 0.524 respectively. Along with that, on the applied frame of 046, the DLADT-PW method has attained a higher accuracy of 0.970 whereas the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.950, 0.853, 0.752, 0.768, and 0.536 correspondingly. Eventually, on the applied frame of 106, the DLADT-PW model has achieved a maximum accuracy of 0.990 but the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF methodologies have led to reduced accuracy of 0.990, 0.912, 0.834, 0.723, and 0.513 respectively. Furthermore, on the applied frame of 136, the DLADT-PW model has obtained a superior accuracy of 0.980 whereas the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.980, 0.946, 0.839, 0.713, and 0.632 correspondingly.

In the same way, on the applied frame of 158, the DLADT-PW model has obtained a maximum accuracy of 0.990 but the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF manners have led to reduced accuracy of 0.990, 0.771, 0.783, 0.716, and 0.544 respectively. Moreover, on the applied frame of 180, the DLADT-PW model has reached a higher accuracy of 0.990 whereas the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.990, 0.853, 0.852, 0.704, and 0.605 correspondingly.

Table 4 has showcased the detection accuracy of anomalies of the DLADT-PW model on the applied test007 video sequence. The resultant table values referred the proficient anomaly detection performance of the DLADT-PW model. For instance, the anomaly 1 in the frames 078, 091, 092, 110, 113, 115, 125, 142, 146, 147, 148, 150, 178, 179, and 180 are detected with the highest accuracy of 0.95, 0.97, 0.99, 0.95, 0.99, 0.99, 0.98, 0.99, 0.98, 0.99, 0.99, 0.96, 0.89, 0.88, and 0.97 correspondingly. Likewise, the anomaly 2 in the frames 078, 091, 092, 110, 113, 115, 125, 142, 146, 147, 148, 150, 178, 179, and 180 are noticed with the superior accuracy of 0.96, 0.99, 0.97, 0.95, 0.83, 0.60, 0.98, 0.95, 0.70, 0.80, 0.60, 0.90, 0.86, 0.78, and 0.88 respectively. Besides, the anomaly 3 in the frames 110, 113, 115, 125, 142, 147, 148, 178, 179, and 180 are noticed with the maximum accuracy of 0.99, 0.99, 0.99, 0.99, 0.97, 0.80, 0.60, 0.65, 0.70, and 0.75 correspondingly.

Table 5 and Fig. 6 demonstrated the comparative outcomes analysis of the DLADT-PW technique with existing techniques on the applied Test007 sequence. The values in the table signified that the SF model has failed to exhibited effective detection performance over all the other techniques. In line with, the MDT and MPPCA models have demonstrated somewhat increased result over the SF model. Concurrently, the Fast R-CNN model has demonstrated moderate outcome whereas a near-optimal detection rate is accomplished by the RS-CNN model. However, the proposed DLADT-PW model has resulted in higher detection performance over all the other compared models. For instance, on the applied frame of 040, the DLADT-PW model has attained a superior accuracy of 0.955 while the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.940, 0.892, 0.842, 0.758, and 0.636 respectively. Likewise, on the applied frame of 046, the DLADT-PW model has obtained a superior accuracy of 0.980 whereas the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.975, 0.928, 0.860, 0.658, and 0.709 respectively. Eventually, on the applied frame of 106, the DLADT-PW manner has obtained a maximum accuracy of 0.860 while the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.833, 0.829, 0.824, 0.704, and 0.652 respectively. Moreover, on the applied frame of

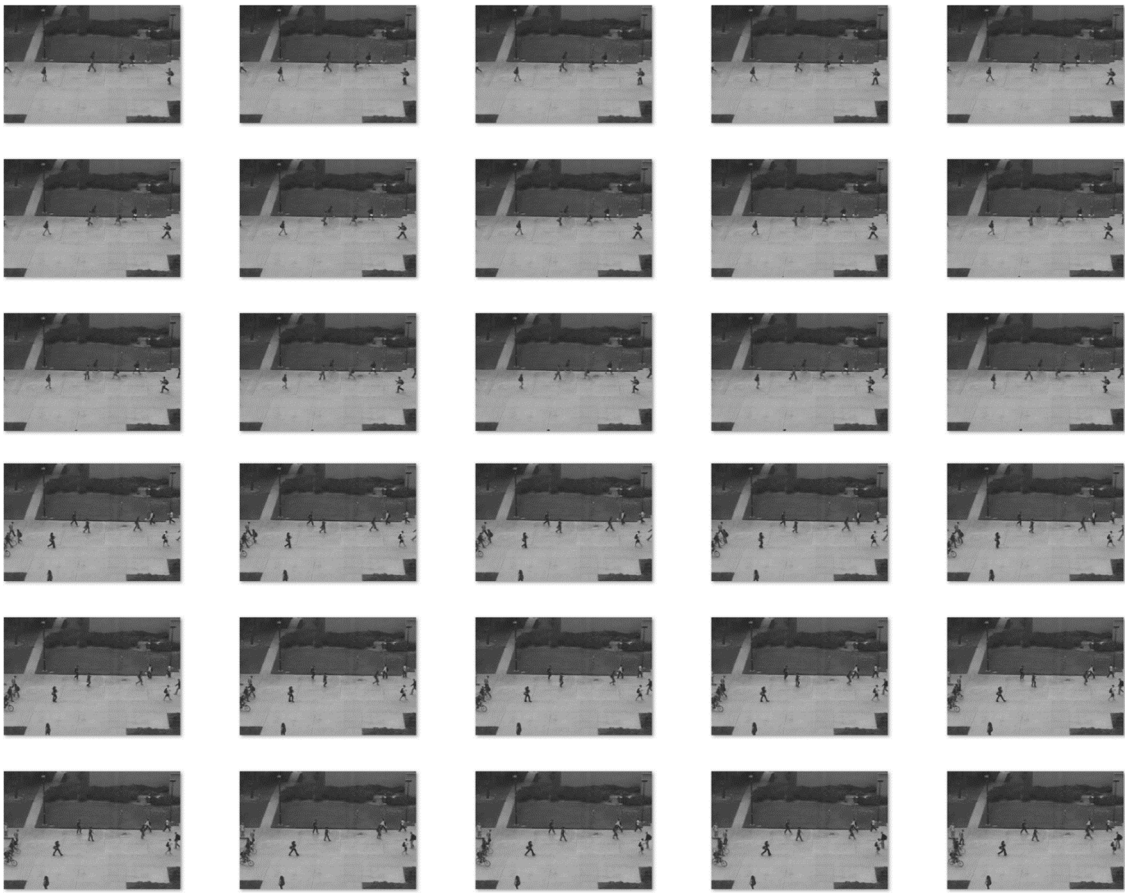

Fig. 3. Sample images.

Fig. 4. (a) Test Image (b) Anomaly Detected Image.

136, the DLADT-PW model has obtained a maximum accuracy of 0.840 whereas the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF methodologies have led to reduced accuracy of 0.835, 0.798, 0.748, 0.788, and 0.687 respectively.

Also, on the applied frame of 158, the DLADT-PW model has achieved a maximum accuracy of 0.930 but the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF approaches have led to reduced accuracy of 0.870, 0.825, 0.811, 0.709, and 0.699 correspondingly. Eventually, on the applied frame of 180, the DLADT-PW model has reached a maximum accuracy of 0.867 whereas the RS-CNN, Fast R-CNN, MDT, MPPCA, and SF models have led to reduced accuracy of 0.830, 0.808, 0.799, 0.769, and 0.705 correspondingly.

Table 6 investigates the average accuracy analysis of the DLADT-PW with existing models on the applied dataset (Shankar and Perumal, 2020). Fig. 7 examines the average accuracy analysis of the DLADT-PW with existing models on the applied Test004 dataset. From the figure, it is evident that the MPPCA and SF models have obtained a reduced average accuracy of 0.746 and 0.564 respectively. Followed by, a slightly improved average accuracy of 0.851 and 0.811 have been obtained by the Fast R-CNN and MDT models. Though the RS-CNN approach has led to a competitive average accuracy of 0.975, the presented DLADT-PW model has accomplished a maximum average accuracy of 0.982.

Fig. 8 observes the average accuracy analysis of the DLADT-PW with existing methods on the applied Test007 dataset. From the figure, it can be evident that the MPPCA and SF models have reached a reduced average accuracy of 0.718 and 0.690 correspondingly. Likewise, a somewhat enhanced average accuracy of 0.821 and 0.778 has been attained by the Fast R-CNN and MDT models. But, the RS-CNN method has led to a competitive average accuracy of 0.867, the proposed

Table 2
Accuracy of anomalies in Test004 Sequences.

Frame Number	Anomaly 1	Anomaly 2
078	0.95	–
091	0.96	–
092	0.97	–
110	0.98	–
113	0.99	–
115	0.99	–
125	0.99	0.97
142	0.99	0.98
146	0.99	0.97
147	0.99	0.99
148	0.99	0.99
150	0.99	0.99
178	0.99	0.99
179	0.99	0.99
180	0.99	0.99

Table 3
Result Analysis of Existing with DLADT-PW model for the test case Test004 in terms of Accuracy.

Frames	DLADT-PW	RS-CNN	Fast R-CNN	MDT	MPPCA	Social Force
040	0.950	0.940	0.819	0.768	0.744	0.524
042	0.960	0.940	0.824	0.766	0.782	0.639
046	0.970	0.950	0.853	0.752	0.768	0.536
051	0.980	0.960	0.793	0.898	0.759	0.601
075	0.990	0.990	0.783	0.827	0.752	0.524
106	0.990	0.990	0.912	0.834	0.723	0.513
123	0.980	0.970	0.913	0.879	0.714	0.575
135	0.985	0.975	0.924	0.806	0.775	0.536
136	0.980	0.980	0.946	0.839	0.713	0.632
137	0.990	0.985	0.917	0.856	0.754	0.579
149	0.990	0.990	0.839	0.786	0.709	0.613
158	0.990	0.990	0.771	0.783	0.716	0.544
177	0.990	0.990	0.793	0.753	0.770	0.522
178	0.990	0.985	0.824	0.765	0.802	0.514
180	0.990	0.990	0.853	0.852	0.704	0.605

DLADT-PW method has accomplished a higher average accuracy of 0.896. From the above-mentioned tables and figures, it is evident that the presented DLADT-PW technique is found to be an effective tool for the detection of anomalies in pedestrian walkways.

5. Conclusion

This paper has designed an automated DLADT-PW model to improve the safety of pedestrians. DLADT-PW model aims to recognize and categorize the dissimilar anomalies present in the pedestrian walkways such as cars, skating, jeep, etc. Firstly, the surveillance video is primarily converted into a set of frames, and anomalies are detected in each frame. Next, the preprocessing is performed to improve the quality of the image. Followed by, the Mask RCNN model is applied for the detection of anomalies and DenseNet 169 model is utilized as the baseline network for Mask RCNN. At last, the anomalies that exist in the frame are

Table 4
Accuracy of anomalies in Test007 Sequences.

Frame Number	Anomaly 1	Anomaly 2	Anomaly 3
078	0.95	0.96	–
091	0.97	0.99	–
092	0.99	0.97	–
110	0.95	0.95	0.99
113	0.99	0.83	0.99
115	0.99	0.60	0.99
125	0.98	0.98	0.99
142	0.99	0.95	0.97
146	0.98	0.70	–
147	0.99	0.80	0.80
148	0.99	0.60	0.60
150	0.96	0.90	–
178	0.89	0.86	0.65
179	0.88	0.78	0.70
180	0.97	0.88	0.75

Table 5
Result Analysis of Existing with DLADT-PW model for the test case Test007 in terms of Accuracy.

Frames	DLADT-PW	RS-CNN	Fast R-CNN	MDT	MPPCA	Social Force
040	0.955	0.940	0.892	0.842	0.758	0.636
042	0.980	0.955	0.918	0.850	0.743	0.723
046	0.980	0.975	0.928	0.860	0.658	0.709
051	0.963	0.947	0.917	0.827	0.710	0.640
075	0.937	0.923	0.877	0.847	0.737	0.665
106	0.860	0.833	0.829	0.824	0.704	0.652
123	0.983	0.980	0.913	0.885	0.686	0.699
135	0.970	0.960	0.942	0.817	0.680	0.724
136	0.840	0.835	0.798	0.748	0.788	0.687
137	0.863	0.827	0.810	0.801	0.673	0.741
149	0.730	0.680	0.546	0.527	0.672	0.633
158	0.930	0.870	0.825	0.811	0.709	0.699
177	0.800	0.733	0.689	0.623	0.756	0.747
178	0.787	0.723	0.629	0.610	0.724	0.689
180	0.867	0.830	0.808	0.799	0.769	0.705

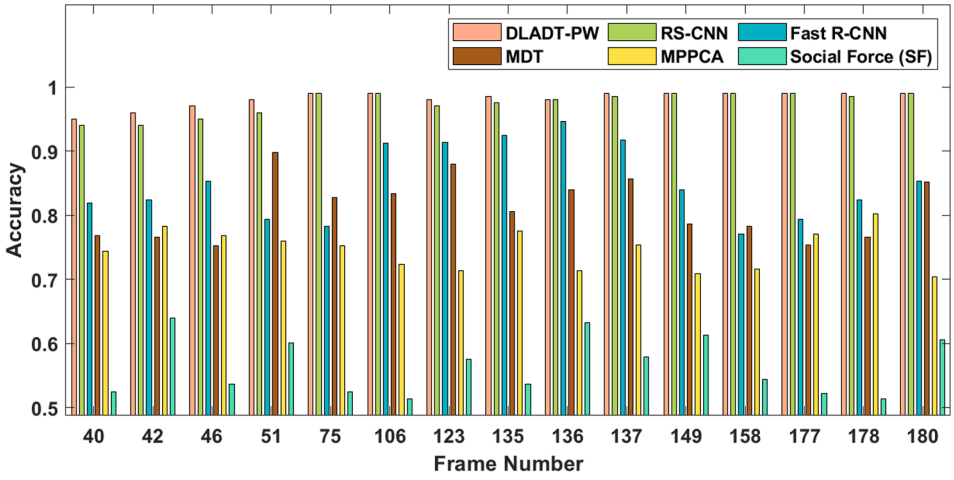

Fig. 5. Result analysis of DLADT-PW model on Test004 dataset.

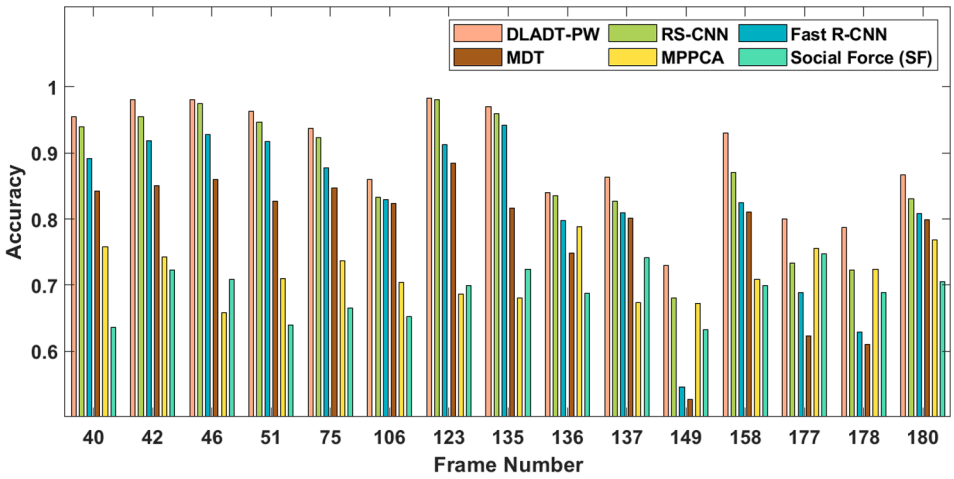

Fig. 6. Result analysis of DLADT-PW model on Test007 dataset.

Table 6
Average Analysis of Accuracy on Applied Dataset.

Methods	DLADT-PW	RS-CNN	Fast R-CNN	MDT	MPPCA	Social Force
Test004	0.982	0.975	0.851	0.811	0.746	0.564
Test007	0.896	0.867	0.821	0.778	0.718	0.690

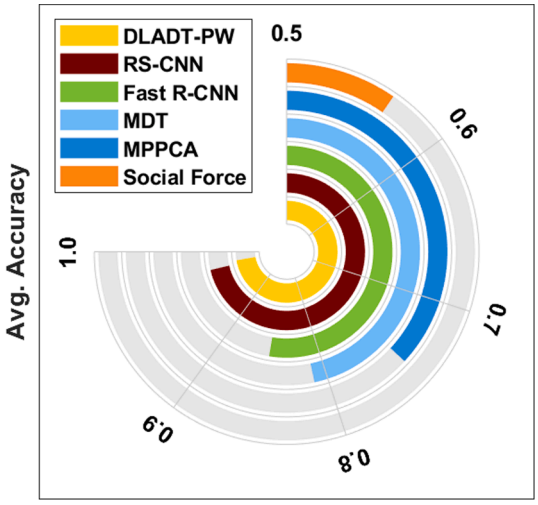

Fig. 7. Average accuracy analysis of DLADT-PW method on Test004 dataset.

successfully identified and classified. For verifying the superior anomaly detection performance of the DLADT-PW technique, a wide set of simulations were accomplished and the results are inspected under distinct aspects. The obtained experimental values confirmed the superior characteristics of the DLADT-PW technique by achieving a maximum detection accuracy. In future work, the presented DLADT-PW technique can be extended to the detection of anomalies under the consideration of poor weather conditions. In the future, the presented work can be implemented in various real-time scenarios like the detection of vehicles in pedestrian walkways. In addition, the presented model can be employed to detect crime scenes like robbery, quarreling, and so on from the surveillance cameras.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

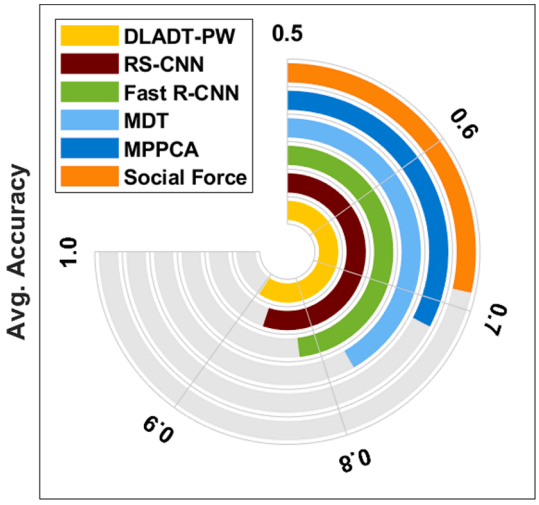

Fig. 8. Average accuracy analysis of DLADT-PW method on Test007 dataset.

Declaration of Competing Interest

The authors declare that they have no conflict of interest. The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

References

Alqaralleh, B.A.Y., Mohanty, S.N., Gupta, D., Khanna, A., Shankar, K., Vaiyapuri, T., 2020. Reliable Multi-Object Tracking Model Using Deep Learning and Energy Efficient Wireless Multimedia Sensor Networks. IEEE Access 8, 213426–213436. https://doi.org/10.1109/ACCESS.2020.3039695.

Cocca, P., Marciano, F., Alberti, M., 2016. Video surveillance systems to enhance occupational safety: A case study. Saf. Sci. 84, 140–148.

Feng, Y., Yuan, Y., Lu, X., 2017. Learning deep event models for crowd anomaly detection. Neurocomputing 219, 548–556.

Gong, Y., Wang, L., Guo, R., Lazebnik, S., 2014. Multiscale orderless pooling of deep convolutional activation features. ECCV 392–407.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm.

Kim, J.; Grauman, K. Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2921–2928.

Kratz, L.; Nishino, K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1446–1453.

Li, Y., Xu, X. and Yuan, C., 2020. Enhanced Mask R-CNN for Chinese Food Image Detection. Mathematical Problems in Engineering, 2020.
Li, W., Mahadevan, V., Vasconcelos, N., 2014. Anomaly detection and localization in crowded scenes. IEEE Trans. Pattern Anal. Mach. Intell. 36, 18–32.
Liu, W.; Luo, W.; Lian, D.; Gao, S. Future frame prediction for anomaly detection—A new baseline. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6536–6545.
Lu, C.; Shi, J.; Jia, J. Abnormal event detection at 150 FPS in MATLAB. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 2720–2727.
Mahadevan, V.; Li, W.; Bhalodia, V.; Vasconcelos, N. Anomaly detection in crowded scenes. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1975–1981.
Murugan, B.S., Elhoseny, M., Shankar, K., Uthayakumar, J., 2019. Region-based scalable smart system for anomaly detection in pedestrian walkways. Comput. Electr. Eng. 75, 146–160.
Rahouti, A., Lovreglio, R., Gwynne, S., Jackson, P., Datoussaïd, S., Hunt, A., 2020. Human behaviour during a healthcare facility evacuation drills: Investigation of pre-evacuation and travel phases. Saf. Sci. 129, 104754.
Ribeiro, M., Lazzaretti, A.E., Lopes, H.S., 2018. A study of deep convolutional auto-encoders for anomaly detection in videos. Pattern Recognit. Lett. 105, 13–22.
Sabokrou, M., Fayyaz, M., Fathy, M., Klette, R., 2017. Deep-cascade: Cascading 3D deep neural networks for nast anomaly detection and localization in crowded scenes. IEEE Trans. Image Process. 26, 1992–2004.
Sabokrou, M., Fayyaz, M., Fathy, M., Moayed, Z., Klette, R., 2018. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Comput. Vis. Image Underst. 172, 88–97.
Shankar, K., Perumal, E., 2020. A novel hand-crafted with deep learning features based fusion model for COVID-19 diagnosis and classification using chest X-ray images. Complex Intell. Syst. https://doi.org/10.1007/s40747-020-00216-6.
Tsai, M.K., 2014. Automatically determining accidental falls in field surveying: A case study of integrating accelerometer determination and image recognition. Saf. Sci. 66, 19–26.
Wester, M., Giesecke, J., 2019. Accepting surveillance–An increased sense of security after terror strikes? Saf. Sci. 120, 383–387.
Xu, C., Wang, G., Yan, S., Yu, J., Zhang, B., Dai, S., Li, Y. and Xu, L., 2020. Fast Vehicle and Pedestrian Detection Using Improved Mask R-CNN. Mathematical Problems in Engineering, 2020.
Xu, D., Song, R., Wu, X., Li, N., Feng, W., Qian, H., 2014. Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts. Neurocomputing 143, 144–152.
Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., Schneider, D., 2020. Automated cattle counting using Mask R-CNN in quadcopter vision system. Comput. Electron. Agric. 171, 105300.
Xu, D., Yan, Y., Ricci, E., Sebe, N., 2017. Detecting anomalous events in videos by learning deep representations of appearance and motion. Comput. Vis. Image Underst. 156, 117–127.
Yang, E., Parvathy, V.S., Selvi, P.P., Shankar, K., Seo, C., Joshi, G.P., Yi, O., 2020. Privacy Preservation in Edge Consumer Electronics by Combining Anomaly Detection with Dynamic Attribute-Based Re-Encryption. Mathematics 8 (11), 1871.
Yu, B., Liu, Y., Sun, Q., 2017. A content-adaptively sparse reconstruction method for abnormal events detection with low-rank property. IEEE Trans. Syst. Man Cybern. Syst. 47, 704–716.
Zhang, Z., Trivedi, C., Liu, X., 2018. Automated detection of grade-crossing-trespassing near misses based on computer vision analysis of surveillance video data. Saf. Sci. 110, 276–285.