ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

The Use of Deep Learning for Pedestrian Path Abnormal Event Detection

[1]Mrs. G. Haritha Rani, [2]Peruru Devi Nagendra Indumathi,

[1] Associate Professor, Department of MCA, Rajamahendri Institute of Engineering & Technology. Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

[2]Student, Department of MCA, Rajamahendri Institute of Engineering & Technology. Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

Abstract—

Because it may pick up on any form of odd movement, a video surveillance system is an important part of the surveillance industry. The security cameras have been used to capture and analyze any unexpected or suspicious actions. Then, we utilize that information to identify patterns that don't fit with typical behavior. Bikers, skateboarders, little carts on pedestrian walkways, etc. are just a few of the ways they appear in film sequences. The primary objective of this research is to design and implement a Video Anomaly Detection (VAD) system that can effectively analyze films for anomalous movements by using deep learning and image processing methods.

Subjects: pattern recognition, deep learning, video analytics, image processing, video surveillance

INTRODUCTION

Concerns over privacy have lately refocused interest in intelligent video surveillance, a complex area of computer vision and machine learning systems. on matters of global safety. Public places such as bus stops, train stations, and shopping malls need heightened vigilance in the event of suspicious behavior. The employment of surveillance cameras has grown in both public and private spaces as a result of technical breakthroughs and lower prices. Human operators are often used to monitor unwanted events. They look at many camera footage shot at the same time. When they go long periods without seeing anything, they lose some of their ability to detect unusual events as they happen. This renders the current setup mostly ineffective outside of forensics as a recording device. As a result, finding and responding to anomalies quickly requires real-time automated anomaly detection. As a result, we're aiming to build an anomaly system that can spot strange movement in videos. Here, the input picture dataset is analyzed using a combination of image processing methods and deep learning techniques, including recurrent neural networks and convolutional neural networks.

LITEARTURE SURVEY

When anything happens that doesn't fit the usual pattern, we call it an anomaly. The most important part of monitoring is identifying anomalies. However, it was consistently determined to be a

problem since there are places where anomalies could be mistakenly found. For instance, whereas shooting a pistol is considered unusual in everyday life, it is considered par for the course at a shooting range. Consequently, certain occurrences are deemed strange even when they conform to local norms. [1][2]Analysis of violent crime using the Waikato Environment for Knowledge Analysis (WEKA) dataset reveals a pattern between actual criminal data and community data. Among the three methods utilized in this study—additional regression, linear regression, and decision stump—linear regression demonstrated the power of deep learning approaches for anomaly identification by determining the event's unpredictability of recurrence. the third A trend was found in the anomaly detection in Philadelphia, according to a research by Kim S et al. [4]. Logistic regression, ordinal regression, decision trees, and k-nearest neighbor are some of the machine learning approaches used to train the model for anomaly detection from vast data sources. A 69% accuracy rate has been reached by the models. [5] Elharrouss et al. [6] utilized a

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

dataset of past crime scenes to make predictions about where crimes are likely to occur. In order to make sense of the data, the levenerg-marquardt method was used. Along with this, the research also made use of the scaled approach to analyze and comprehend the data. The most effective method, which made use of a scaled approach, had an accuracy rate of 78%. As an added bonus, it may reduce crime by as much as 78%. [7] [8] With data integrated into a 200x250 m grid, Sultani et al. performed a comprehensive analysis on anomaly detection in urban regions. It was evidently evaluated in retrospect. Anomaly detection methods like neural networks and ensemble logistic regression were part of the approach they suggested. Based on the findings, it is more accurate to do anomaly prediction every 14 days rather than monthly. [9] Rummens et al. conducted an extensive analysis of anomalous activity using anomaly data from the last fifteen years, beginning just before 2017. For the detection, methods like decision trees and k-nearest neighbors have been used. In testing on a dataset of 5,60,000 anomalous activity, it achieved an accuracy level of 39-44%. [10] the eleventh

METHODOLOGY

In deep learning, a kind of machine learning, there are extra levels beyond the input and output layers known as hidden layers. With deep learning, it's possible to reproduce

any autonomous human activity. Adding additional hidden layers improves the model's accuracy. Even with a single layer, a deep learning model won't provide the results you're hoping for. [12] Here, the model was built using RNN and CNN, two deep learning approaches. The model receives a large quantity of images as input. Several security cameras in various locations captured these images. By using various image processing methods, the collection's images are examined. The model's structure allows it to evaluate and identify patterns. In the first step, the picture dataset is divided into two parts: the train set and the test set. The train data is used to train the model, and the test data is used to evaluate its performance. Models have been constructed using deep learning approaches, such as convolutional neural networks (CNNs). Each algorithm has been evaluated for its accuracy, performance, and other characteristics; the model-building process then employs the method with the best accuracy. in [13] [14]

IV. SYSTEM ARCHITECTURE

The system architecture is presented in fig.1.

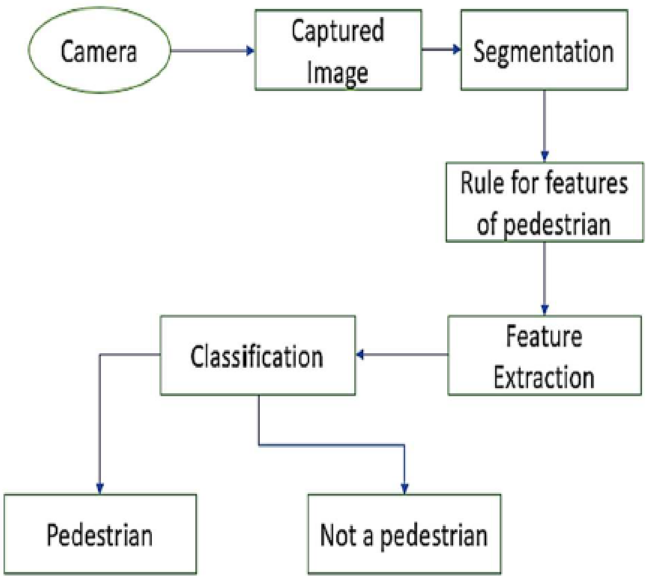


Fig.1. System architecture

Picture, audio, and video are only a few of the input data formats. To avoid duplication, we utilized pedestrian photo data that had already been analyzed. The next day,

The method of feature extraction is used. At this stage, the essential attributes are extracted from the image. The next step is to implement the Feature Selection. In this step, we use the features acquired in the previous step to choose the most relevant attributes. There should now be two sets of data: one for testing and one for training. Next, the Deep Learning model is fine-tuned to predict whether the input image has an out-of-the-ordinary object. [15] section of the sidewalk Included in the dataset are two folders: ped1 and ped2. People are shown walking in both directions in ped1, which consists of 34 training video samples, and ped2, which consists of 36 testing video samples, 16 training video samples, and 12 testing films. Figures 2 and 3 display excerpts from the video dataset. [17]

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Fig.2. Captured image

Fig.3. Captured image

DATA PREPROCESSING

Anomalies may be detected by observing its motion, which includes both the object's speed and its appearance. Object detection and optical flow are two methods that can help with this. Utilizing these techniques

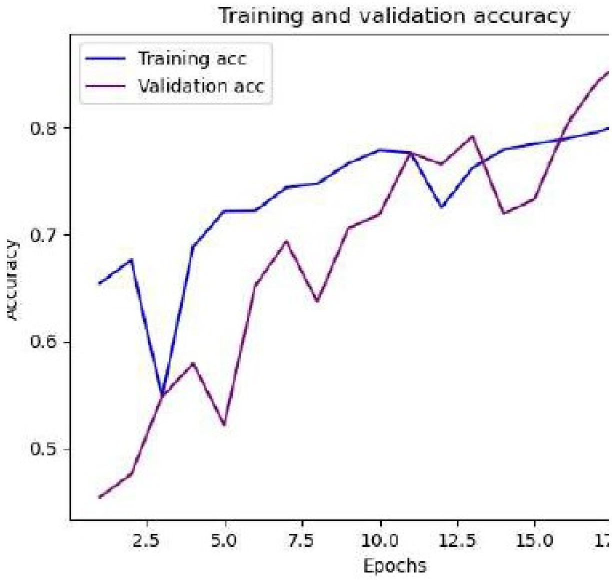
to the video frames included in the Object Detection dataset.: To recognize the item in the video frame, object detection is essential. It lets us identify things from the frames of the video, such as people or anomalies like motorcycles, automobiles, trucks, skaters, etc. Optical Mobility: Optical flow is a method for determining the mobility of objects in a movie by comparing successive frames. Using the brightness constancy assumption, which is dependent on the pixel's brightness, it may detect anomalies, or items moving quickly, in the pedestrian track. [18] in [19]
BATCH ESTABLISHMENT
We use the batch normalization method to accelerate the learning of the model in this project. If this layer works as promised, it should make our model easier to comprehend and the input image training process go by faster. When we choose the batch size of 32 in this model, we send 32 picture samples via the network at once. We use a batch normalization layer to expedite the network's learning process using the training data. In [20],
B. EVALUATION AND EDUCATION
Different camera positions are used to create the "ped1" and "ped2" folders inside the dataset, which are then used to train and test the model. There are a total of 36 video clips for Pred1's training and 34 for Pred2's testing, with 16 for the former and 12 for the latter. Every video clip is reduced to 200 frames so that the models may be trained. Pred1 and pred2 both feature binary flags that distinguish between normal and abnormal. You may also use the given masks to assess the stats. For optimal model performance, we choose the sets according to their sizes; in particular, the training data should be larger than the testing data. [21]
The ability to forecast output is a necessary feature of any model training on input data. The next step is to get the model ready for training, which involves practicing the recovered pedestrian characteristics from the input picture using the Epoch technique. The "epoch" is the number of cycles that should iterate across the training set. As seen in figure 4, this model was trained using 20 epochs, which translates to 20 feeds of the training data into the model. We retrained the data using deep learning algorithms 20 times. Every time, we check the accuracy. The trained model may then be tested. In this scenario, loading the concealed image is necessary for result forecasting. [22] is a

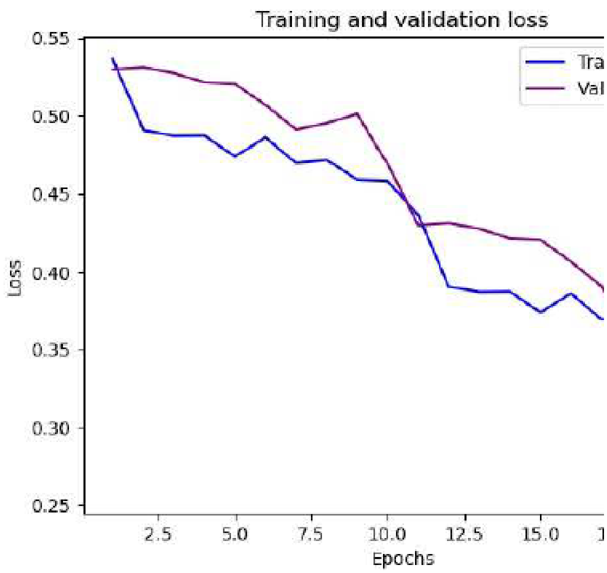
```
Epoch: 1/20..  Training Loss: 0.468..  Validation Loss: 0.597..  Validation Accuracy: 0.675
Epoch: 2/20..  Training Loss: 0.378..  Validation Loss: 0.529..  Validation Accuracy: 0.700
Epoch: 2/20..  Training Loss: 0.349..  Validation Loss: 0.543..  Validation Accuracy: 0.738
Epoch: 3/20..  Training Loss: 0.335..  Validation Loss: 0.481..  Validation Accuracy: 0.750
Epoch: 4/20..  Training Loss: 0.328..  Validation Loss: 0.534..  Validation Accuracy: 0.744
Epoch: 4/20..  Training Loss: 0.295..  Validation Loss: 0.545..  Validation Accuracy: 0.744
Epoch: 5/20..  Training Loss: 0.283..  Validation Loss: 0.679..  Validation Accuracy: 0.700
Epoch: 6/20..  Training Loss: 0.326..  Validation Loss: 0.528..  Validation Accuracy: 0.731
Epoch: 6/20..  Training Loss: 0.290..  Validation Loss: 0.513..  Validation Accuracy: 0.769
Epoch: 7/20..  Training Loss: 0.292..  Validation Loss: 0.503..  Validation Accuracy: 0.781
Epoch: 8/20..  Training Loss: 0.277..  Validation Loss: 0.534..  Validation Accuracy: 0.769
Epoch: 8/20..  Training Loss: 0.285..  Validation Loss: 0.447..  Validation Accuracy: 0.769
Epoch: 9/20..  Training Loss: 0.284..  Validation Loss: 0.551..  Validation Accuracy: 0.725
Epoch: 10/20..  Training Loss: 0.270..  Validation Loss: 0.520..  Validation Accuracy: 0.800
Epoch: 10/20..  Training Loss: 0.276..  Validation Loss: 0.407..  Validation Accuracy: 0.806
Epoch: 11/20..  Training Loss: 0.253..  Validation Loss: 0.484..  Validation Accuracy: 0.812
Epoch: 12/20..  Training Loss: 0.265..  Validation Loss: 0.592..  Validation Accuracy: 0.750
Epoch: 12/20..  Training Loss: 0.254..  Validation Loss: 0.596..  Validation Accuracy: 0.787
Epoch: 13/20..  Training Loss: 0.252..  Validation Loss: 0.462..  Validation Accuracy: 0.794
Epoch: 14/20..  Training Loss: 0.238..  Validation Loss: 0.519..  Validation Accuracy: 0.787
Epoch: 14/20..  Training Loss: 0.244..  Validation Loss: 0.432..  Validation Accuracy: 0.806
Epoch: 15/20..  Training Loss: 0.222..  Validation Loss: 0.503..  Validation Accuracy: 0.806
Epoch: 16/20..  Training Loss: 0.238..  Validation Loss: 0.475..  Validation Accuracy: 0.787
Epoch: 16/20..  Training Loss: 0.271..  Validation Loss: 0.419..  Validation Accuracy: 0.806
```

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Fig. 5 and fig 6 show that the accuracy and loss graph forthe Convolution neural network (CNN). In this project, we

train 20 epoch in the training phase of the model.





MODEL SELECTION

The many-to-many LSTM and convolution neural networks are two types of recurrent neural networks. In this case, deep learning methods are used.

The neural network's backpropagation aids in model building, and the learning rates may be used for improved model training, all because of the network's capacity to extract features, train them with several layers, and achieve greater performance.

CNN, or Convolutional Neural Net

We use a deep learning technique similar to CNN in this study. It is not an easy undertaking to construct a CNN pedestrian anomaly detection model.

In order to construct deep learning models more effectively, convolutional neural networks (CNNs) include the following layers.

KERNAL

Using the Kernal method, we may extract the information or attributes from the camera-captured input photographs and store them as a matrix of pixels. This project's 3x3 kernel size indicates that the size of the 3x3 matrix is determined by comparing the picture's input pixel matrix with the kernel matrix, which is likewise 3x3. A feature extraction matrix is the result of multiplying the two matrices. Given that we previously extended the model, the kernel size is 3x3. [23] The

A smooth surface

After the convolution layer, a flattening layer may be added to reduce the multi-dimensional array of image attributes in this project to a single dimension. This reduced array of image properties is sent into the fully connected network as an input layer, which then utilizes it to decide if the given picture is malignant or not. Consequently, the network's overall performance and processing speed would be enhanced by adding this layer after the convolution layer and lowering the size of the features map. [24]

Conversely, we check whether the picture is out of the ordinary using the Sigmoid function. The sigmoid function takes values between zero and one. The network becomes non-linear because to this sigmoid function, which allows it to learn complex models. Hence, the rectified Linear unit function and the Sigmoid function are both used in this project. This model auto-refines the neural network's parameters using the Adaptive Moment Estimator (Adam), which estimates the values from the network's previous layer of neurons. Both the learning rate and the accuracy of the model are enhanced with the help of this optimizer during sample training. [25] The [26] in

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Layer (type)	Output Shape
conv2d (Conv2D)	(None, 254, 254, 16)
max_pooling2d (MaxPooling2D)	(None, 127, 127, 16)
conv2d_1 (Conv2D)	(None, 125, 125, 32)
max_pooling2d_1 (MaxPooling 2D)	(None, 62, 62, 32)
conv2d_2 (Conv2D)	(None, 60, 60, 16)
max_pooling2d_2 (MaxPooling 2D)	(None, 30, 30, 16)
flatten (Flatten)	(None, 14400)
dense (Dense)	(None, 256)
dense_1 (Dense)	(None, 1)

Fig.8. Predicting model

In this research, we train a model to detect abnormalities in pedestrian images and then use that information to make predictions. All of them are evaluated using the one with the greatest accuracy score.

based on the ultimate accuracy score. This project's accuracy is provided by a Convolution Neural Network (CNN) at 80%.

CONCLUSION

Due diligence was conducted on the anomaly since the well-being of the people must be prioritized above all else, and any potential source of discomfort must be investigated.

In the context of pedestrian pathways, these anomalies may take the form of anything out of the ordinary, such as motorcycles, trucks, skatters, etc. In order to identify these potential hazards to pedestrians, this code employs deep learning and image processing techniques to break down video into frames. Afterwards, the frames are preprocessed to extract useful features from the dataset, and any noise is removed. Then, it uses motion detection and object detection to classify objects based on their size and how they're moving, excluding pedestrians. The video frames serve as input for this model, which may have its learning rate tweaked for faster learning. To get the most out of the model, we additionally tweak its hyperparameters. By utilizing deep learning techniques like recurrent neural networks and convolutional neural networks, which have multiple hidden layers, it is possible to achieve high accuracy and better feature extraction. Consequently, our model is a top-notch performer when it comes to predicting route abnormalities, with an accuracy rate of 80%. Then, by spotting the outliers, we can improve the safety of the route users. The model is trained using object and motion detection. The model may be improved by adding additional parameters to this. This model has the potential to be improved and provide more accurate results with the use of different parameters.

REFERENCES

[1]. "Evaluating the use of public surveillance cameras for crimecontrol and prevention." by Vigne, N.G.L., Lowry, S.S.,Markman, J.A., Dwyer, A.M. in US Department of Justice,Office of Community Oriented Policing Services. UrbanInstitute, Justice Policy Center, Washington, DC (2011)

[2]. "A world with a billion cameras watching you is just aroundthe corner." by Lin, L., Purnell, N. in Wall Str. J. (2019)

[3]. "spatio-temporal adversarial networks for abnormal eventdetection." by Lee, S., Kim, H. G., Ro, Y. M.: STAN inIEEE international conference on acoustics, speech andsignal processing (ICASSP)(2018)

[4]. "Observe locally, infer globally a space-time MRF fordetecting abnormal activities with incremental updates." ByKim, J., Grauman, K. in IEEE Conference on ComputerVision and Pattern Recognition (2009)

[5]. "A review of video surveillance systems" by O. Elharrouss, N.Almaadeed, S. Al-Maadeed in J. Vis. Commun. ImageRepresent. (2021)
[6]. "E2E-VSDL: end-to-end video surveillance-based deeplearning model to detect and prevent criminal activities" byM.Q. Gandapur in Image Vis Comput. (2022)
[7]. "Survey on contemporary remote surveillance systems forpublic safety" by

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

T.D. Raty in IEEE Transactions onSystems, Man and Cybernetics. (2010)

[8]. ”Abnormal event detection at 150 FPS in MATLAB.” by LuC., Shi, J., Jia, J. in IEEE International Conference onComputer Vision (2013)

[9]. “A unified approach to interpreting model predictions.”byLundberg, S.M., Lee, S. In Proceedings of the 31stInternational Conference on Neural Information ProcessingSystems(2017)

[10]. “Robust real-time unusual event detection using multiplefixed-location monitors.” by Adam, A., Rivlin, E.,Shimshoni, I., Reinitz, D. in IEEE Trans. Pattern Anal.Mach. Intell. (2008)

[11]. “Video anomaly detection with compact feature sets foronline performance.” by Leyva, R., Sanchez, V., Li, C.-T.in IEEE Trans. Image Process.(2017)

[12]. “Real-world anomaly detection in surveillance videos” by W.Sultani, C. Chen, M. Shah in Proceedings of the IEEEConference on Computer Vision and Pattern Recognition.(2018)

[13]. Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L.,Regazzoni, C., Sebe, N.: Abnormal event detection invideos using generative adversarial nets. In: 2017 IEEEinternational conference on image processing (ICIP), pp.1577–1581 (2017)

[14]. “Crime analysis through machine learning” by S. Kim, P.Joshi, P.S. Kalsi, P. Taheri in IEEE 9th Annual InformaationTechnology, Electronics and Mobile CommunicationConference(IEMCON),IEEE (2018)

[15]. “Decision-theoretic saliency: computational principles,biological plausibility, and implications for

[16]. neurophysiology and psychophysics.” by Gao, D.,Vasconcelos, N. in Neural Comput.(2009)

[17]. “A causal framework to determine the effectiveness ofdynamic quarantine policy to mitigate COVID-19.” ByKristjanpoller, W., Michell, K., Minutolo, M.C. in Appl.Soft Comput. (2021)

[18]. “Hierarchical context modeling for video event recognition.”by Wang, X., Ji, Q. in IEEE Trans. Pattern Anal. Mach.Intell.(2017)

[19]. “Detecting anomalies in image classification by means ofsemantic relationships.” by Pasini, A., Baralis, E. in IEEESecond International Conference on Artificial Intelligenceand Knowledge Engineering (AIKE), Sardinia, Italy.(2019)

[20]. “The use of predictive analysis in spatiotemporal crimeforecasting: building and testing a model in an urbancontext” by A. Rummens, W. Hardyns, L. Pauwels in Appl.Geogr. (2017)

[21]. “Segmentation of COVID-19 pneumonia lesions: a deeplearning approach.” by Ghomi, Z., Mirshahi, R., Bagheri,A.K., Fattahpour, A., Mohammadiun, S., Gharahbagh,A.A., Djavadifar, A., Arabalibeik, H., Sadiq, R., Hewage,K. in Med. J. Islam. Repub. Iran(2020)

[22]. [21] “SiTGRU: single-tunnelled gated recurrent unit forabnormality detection” by H. Fanta, Z. Shao, L. Ma. in Inf.Sci. (2020)