Pakistan Journal of Engineering and Technology
ISSN (Print): 2664-2042; ISSN (Online): 2664-2050
Received: 2024-09-03
Accepted: 2024-12-26
https://doi.org/10.51846/vol7iss4pp166-172

PAKJET
Pakistan Journal of Engineering & Technology

Article

Engineering Smart Detection Systems: Leveraging Neural Networks for Crime Prevention

Imran Hussain [1], Ammara Alvi [1], Sajjad Hussain [1], Danish Ghaffar [2,*], and Mudasira Khalil [3]

[1] Department of Computer Science, The Islamia University of Bahawalpur, Rahim Yar Khan 64200, Pakistan
[2] Institute of Computer Science, Virtual University of Pakistan, Rahim Yar Khan 64200, Pakistan
[3] Institute of Computer Science, KFUEIT, RYK, Rahim Yar Khan 64200, Pakistan
* Correspondence: Danish Ghaffar (danishghaffar443@outlook.com)

Abstract: In recent years, numerous techniques for surveillance security have emerged to enhance public safety and mitigate the risks associated with deviant human activities. This research addresses deviant activity detection by proposing an advanced surveillance system capable of identifying actions such as smoking, harassment, and fighting. The system employs a hierarchical technique and utilizes Convolutional Neural Networks (CNN) to analyze captured images and detect deviant behaviors. A diverse dataset was compiled and augmented with techniques like rotation, zoom, and horizontal flip to improve model robustness and performance. The proposed model achieved an accuracy of 93.33% in identifying deviant classes during testing and validation. The system significantly reduces the need for human intervention by automating the detection process, thereby enhancing response times and overall security. Extensive empirical observations validate the effectiveness of the system in various environments and conditions.

Keywords: Deviant Activity, Image Processing, Security system, CCTV Camera, Auto-Detection

1. Introduction

The investigation of illegal activities and strange occurrences may be exceedingly challenging. The ability to anticipate the layout of a criminal investigation may make the work of law enforcement authorities easier. It is estimated that by the end of 2019, 2500 petabytes of data per day will be generated by the world's CCTV cameras [1], up from 566 petabytes in 2015. It is necessary to have a computer-controlled, effective online surveillance system that can control the operation and identify deviant or inexplicable behavior of items and humans from video in a fashion that is as close to real-time as possible. This is because regular testing of such video content is far beyond the capabilities of video controller professionals. The National Retail Security Survey (NRSS) [2] estimates that in 2019, the retail industry in the United States lost $61.7 billion caused by inventory decline (loss of merchandise due to theft shoplifting, mistake, or fraud. Daily, people fall victim to a wide variety of frauds from diversions and bar code swapping to booster packs and phony weight techniques no one can possibly keep track of them all. Overwhelming, from the perspective of monitoring. Extremely large numbers of video feeds are produced by surveillance camera systems, and the monitoring crew simply cannot analyze them quickly enough. With an increasing number of options for capturing, keeping tabs on all of them might become a daunting effort. Because of its vast application for tracking public and confidential locations, preparedness, aged medical systems, defense systems, and transport systems, automated optical surveillance has evolved into one of the highly prized subjects under study in academics and commercial enterprises. As a result, deploying Closed Circuit Television (CCTV) cameras has become a popular way to track continuous operations and ensure worldwide protection. The worldwide CCTV surveillance market is expected to develop at a compound annual growth of 16.6% from 2017 to 2025, owing to lower costs, simplicity of use, and the customizable layout of sensors [3]. Human constraints have made a real-time evaluation of security cameras a laborious operation. Visual Focus of Attention (VFOA) is the basic human constraint [2]. The human sight could only focus on one location at a time. Even with enormous displays and elevated lenses, a human can only focus on a limited portion of the picture at a moment. There is an espionage situation, the optical focus is a substantial human-related drawback. A violation might occur on a separate display segment or on a distinct display and go unnoticed by the crew. Additional substantial challenges may be connected to, among other things, concentration, monotony, diversions, and a lack of expertise [4], [5].

The rise in deviant and criminal activities has necessitated the development of advanced surveillance systems. Traditional methods often rely heavily on human intervention, which can be inefficient and error-prone. This research focuses on creating a system that minimizes human involvement by leveraging artificial intelligence, specifically CNNs, to detect and classify deviant activities. By automating the detection process, the system aims to provide real-time alerts and responses to potential threats, thereby enhancing overall security.

The fundamental aspects of the system are defined as follows: unusual activity definition (smoking, harassment and fighting), background subtraction, object recognition, tracking of activities, and activity categorization. The data augmentation method is used to describe the activities that are deemed dubious. In order to identify the behavior, motion characteristics that vary between the two or various objects are retrieved.

1.1. Objectives

- Detect deviant activity
- Enhance the security of public surveillance
- Minimize Human Intervention
- Reduce noise and improve accuracy
- Recognize moving objects

1.2. Research Question

There are numerous challenges in this project; these challenges arise due to changes in environmental conditions such as lighting, reflections, and shadows; thus, it is a complex issue that must be dealt with effectively by employing a robust surveillance system. We reduced the issue by employing image rescaling, rotation, zooming, and flip image operations. Our model in training takes different images in testing and validation; therefore, results are changed when rerun, or our model for training and testing.

2. Literature Review

A real-time deviant activity identification study [6] Anomaly detection varies from classification: 1) Negative examples are hard to list. 2) Scarcity makes negative sample collection difficult. Using videos of usual occurrences as training data to construct a model and identify aberrant events is a frequent anomaly detection strategy. Reduced depth sensors are indoor-only and offer low-resolution and noisy depth information, making human posture estimation from depth photographs problematic. Neural networks will resolve these concerns. Most anomaly detection research has utilized unsupervised learning, while their experiment employed supervised learning. Long short-term memory, AMD algorithm, SURF, convolutional neural network (ANN), Gaussian Classifiers, and sparse autoencoder are machine learning and deep learning methods. Initially, data is collected from several sites and social media applications based on specific characteristics.

Another study aims to recognize abnormal behavior, such as cheating and misconduct during exams. The system collects crucial frames from a video series centered on human action, and uses a deep learning model 2D and 3D CNN for classifying and identifying malicious actions. Three datasets were employed for empirical study in order to evaluate this system. That is an original dataset called Examination Unusual Activity (EUA), followed by two more typical datasets called Violent Flow and Movies. In the form of educational assessment invigilation, there really is no conventional dataset for identifying malicious actions, so, they created their own dataset to evaluate their system. This technique is developed for recognizing unusual activity in academic environments. Examination Unusual Activity The dataset was produced to categorize the actions as normal or aberrant. Three actions have been used to identify malicious inactivity of cheating behavior: head movement, material transfer, and communication [7]. To detect and differentiate the normal and anomalous activities Spatial-temporal approach of deep learning is used. It is influenced by the notion of rote memorization and the continual learning process in human cognition. The proposed technique is evaluated using three baseline datasets: the CUHK Avenue dataset, the UCSD Pedestrian 1 and UCSD Pedestrian 2 datasets, and the CUHK Avenue dataset. Using this empirical survey, they show that the ISTL model can detect and localize anomalies in a near-sufficient manner and that it outperforms state-of-the-art abnormality detection methods published in the literature [8]. Reference [9] explains the methods used to identify "click-on fraud." To perfect online target advertising, researchers looked at several different machine-learning approaches to advertising methods used online. User-centric approaches and those that are content-centric are the two primary types that are defined and categorized when it comes to targeted internet advertising strategies and related machine learning-based procedures. In addition to this, the paper notes the need for click-fraud detection systems that are based on machine learning to safeguard advertisers against fraudulent clicks. This classification lays the groundwork for future research that will examine the techniques of machine learning and artificial intelligence to further optimize the targeting and effectiveness of online advertising strategies, as well as to address security and privacy concerns in advertising, such as the detection of click fraud [10]. They explain a semantic approach to advertising suggestions in this study that they carried out. The framework that is suggested in this research makes use of ontologies to semantically represent both the interests of users and the primary characteristics of adverts. The usage of this common model makes it much easier to uncover meaningful matches between users and advertisements, which increases the likelihood that the advertisements will be appealing to the users. The framework, in its current form, makes use of natural language processing (NLP) tools to automatically process textual content that is concerned with users as well as ads, and it then generates vectors that represent user profiles as well as advertisement profiles following the domain ontology. The ad recommender framework has undergone rigorous validation in a simulated environment, resulting in an aggregated f-measure of 79.2% and a Mean Average Precision at 3 (MAP@3) of 85.6%. The method known as LSTM was used by the authors of this study to provide consumers with an advertisement display. In this research, the authors center their attention on the challenge of developing a novel framework employing LSTM-based deep neural networks to predict user click responses and user interests. Their method permits sequences to have variable lengths and different numbers of dimensions, and it can maximally leverage the temporal information contained in user sequences for learning. This is accomplished through the use of padding and bucketing to learn binary user

click prediction and multi-class user interest prediction. Experiments and comparisons on real-world data collected from our industry partner show that the method can encode useful latent temporal information in request sequences to predict users' responses and interest in online digital advertising. The data used for these experiments and comparisons came from an industry partner [11].

This study gives a state-of-the-art assessment of the many methodologies of community recommendation systems that are currently in use. The purpose of this survey is to investigate social media, the various forms it might take, and the part it plays in the recommender system. They carry out a comprehensive analysis of the various SRS formats by making use of a variety of research methods. In addition to this, they talk about the many strategies, variables, and datasets that were used in the construction of distinct SRS. Using the study articles that were collected for the assessment, they determine the various fields in which investigations are carried out. Using the criteria that they developed, they make a comparison between the various research publications. Additionally offered are a variety of datasets that were utilized in a variety of SRS [12].

2.1. Contribution

- Developing an advanced surveillance security system using CNNs for detecting deviant activities.
- Implementing a hierarchical technique for identifying multiple deviant behaviors.
- Compiling and augmenting a diverse dataset to enhance model robustness.
- Achieving a high accuracy rate of 93.33% in predicting deviant classes.
- Reducing human intervention by automating the detection process.
- Validating the system's performance through extensive empirical observations.

3. Materials and Methods

The methodology is explained in the flow chart shown in Fig. 1.

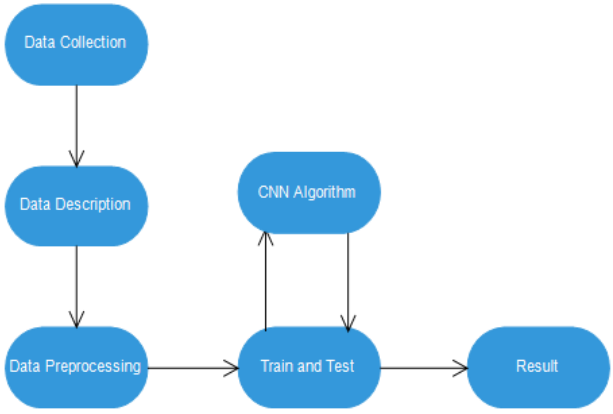

Figure 1. Flow chart to explain the methodology.

3.1. Data Collection

We have collected different images of human deviant activity. These images belong to different classes, like smoking, harassment, and fighting. We use different sources to collect images, like CCTV cameras images, Google images, etc. The main purpose of collecting different images from different to train our model in different environments of images. These different images have different backgrounds and many human actions related to deviant classes. Different foregrounds and backgrounds have improved the performance of our model. In any scenario, the model predicts accurately and gives a good result.

3.1.1. Data Description

We use three classes of Deviant activity smoking, harassment, and fighting. All these images are in a jpg extension file. All images have an RGB image. Fig. 2 shows the overall distribution of dataset.

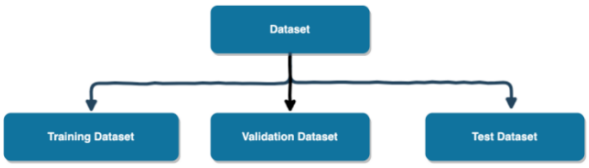

Figure 2. Dataset distribution.

3.1.2. Data Preprocessing

The dataset must first be loaded, and then the unique label class must be examined. The next step is to investigate the training dataset, during which you should verify both the overall number of photos and each individual training dataset class. After verifying the total number of photos included inside the train classes, the next step is to examine the Data Frame, which contains the file paths in one column and the class name in the second column. The images should be loaded using an image data generator in conjunction with data augmentation. The goal of data augmentation is to artificially enhance the total quantity of data that is accessible, and it consists of a variety of techniques for extracting new data components from existing data sets. You have the ability to rotate pictures freely between 0 and 360 degrees if you provide the Image Data Generator class with an integer value to use for the rotation range parameter.

Image Data Generator was another tool that helped us train our model in a more effective manner. In order to train our model to work with images of varying sizes and shapes, I make use of a variety of properties, such as rotation, zoom, width and height range, shear range, and horizontal flip. Because every picture has a lot of color, I also designate the image mode as "RGB."

3.1.3. Training and Testing

To train to detect deviant activity we have used the deep learning algorithm Convolutional Neural Networks CNN. First of all, we know about deep learning and its algorithm and why we use deep learning techniques in our thesis.

3.1.4. Data Augmentation and Image Generator

The images should be loaded using an image data generator in conjunction with data augmentation. The goal of data augmentation is to artificially enhance the total quantity of data that is accessible, and it consists of a variety of techniques for extracting new data components from existing data sets. Examples of this include making minor alterations to the data or making use of deep learning models in order to create extra data points. By adding fresh and original examples to existing datasets for training purposes, data augmentation may help improve the effectiveness of machine learning algorithms and the results they produce. When the dataset is complete and contains enough information, a machine learning model performs significantly better and more accurately.

When developing machine learning models, data collecting and labeling may be a time-consuming and expensive process. Through the use of various data augmentation strategies, businesses are able to reduce the costs associated with their operating overheads. Image Data Generator was another tool that helped us train our model in a more effective manner. Image enhancement function inside Keras that is accessed through Image Data Generator.

3.2. Training Model using CNN

Now that CNN has been used to train our model, I can move on. Due to its ability to handle enormous amounts of data, deep learning has recently shown to be a very valuable technology. The use of traditional approaches has been overtaken in popularity by the use of hidden layers, notably in pattern recognition. Convolutional Neural Networks are among the deep neural networks that are utilized the most often nowadays.

During the process of our model training, I have utilized the use of the following CNN function:

- Layers
- An activation function comparable to Rely and Soft Max
- Optimizer Adam
- Batch size
- Epochs

3.2.1. Layers

If I have an input that is of size A x A x D and a D out number of kernels with a spatial dimension of L, stride N, and amount of padding M, then I can use the following formula to compute the size of the volume that will be output if I have an input that is of size A x A x D.

$$\mathrm{A_{out}} = \frac{\mathrm{A - L + 2M}}{\mathrm{N}} + 1 \qquad (1)$$

It is the responsibility of the Pooling layer, which may also be referred to as the Convolutional Layer, to bring down the overall dimensionality of the Convolved Component. By reducing the size, the amount of computing power needed to analyze the data may be cut down significantly. Pooling may be broken down into two categories: maximum pooling and average pooling. If I have an activation map with the dimensions A x A x D, a pooling kernel with the dimensions L x N, and stride N, then I can use the following formula to compute how large the output volume will be.

$$A_{out} = \frac{A - L}{N} + 1 \qquad (2)$$

3.2.2. Activation Function

It is essential for a neural network to have a good activation parameter. A basic linear regression model that does not include an activation function is what a neural network is. This demonstrates that the activation parameter is responsible for providing non-linearity in neural networks. An activation function is nothing more than a straightforward mechanism that transforms its inputs into outputs that fall within a predetermined value range. There are many different kinds of activation functions, and each one works in a somewhat different manner to achieve the same goal. If the input is positive, the rectified linear activation function (also known as ReLU) is a nonlinear linear function that produces the same value as the input directly. If the input is negative, the function outputs zero. It has become the standard activation function for a number of different kinds of neural networks due to the fact that models that make use of it need less time to be trained and, on average, generate better results. The following equation describes how this activation function works:

$$f\,(x) = \max(0,\ x) \qquad (3)$$

SoftMax could be a scientific work that changes a vector of integrality into a vector of probabilities, with the likelihood of each esteem relative to the vector's relative scale.

3.2.3. Optimizer Adam

In order to incrementally improve the network weight based on the training data, I employed the Adam optimization technique rather than the more conventional stochastic gradient descent process. Adam is an optimization strategy that was developed by combining the most beneficial aspects of the AdaGrad and RMSProp techniques. This strategy was developed for noisy problems that have sparse gradients.

3.2.4. Batch Size and Epochs

The number of samples that are sent to the network all at once is referred to as the batch size. A hyper-parameter known as the batch size determines how many samples must be processed before the internal model parameters may be changed. The sample supplies the inputs that the algorithm needs, as well as an output that can be compared to the estimate in order to determine an error. The number of times that the having-to-learn method is executed across the whole training dataset is controlled by a hyper-parameter called the frequency of epochs. Every single sample that was used for the training dataset was given the opportunity, once for each epoch, to make adjustments to the fundamental model parameters. An era may include a single group or many groups.

4. Results

After I have successfully trained the model, the next step is to see and assess the accuracy of the results. Our model has a value accuracy of 82% (see Fig. 3). Accuracy is defined as the degree to which the standard against which one was taught is right. "Value accuracy" was the validation set that was used. The value accuracy parameter refers to a set of samples that were not presented to the network during training and so provides an overall evaluation of how well your model performs in situations that are not part of the training set.

The value loss cost function value is used for the cross-validation of the data, and the training data is also loss-processed using the cost value function. Neurons that utilize fallout won't lose any randomized neurons' data throughout the validation process. The reason for this is that during training, I make use of dropout to provide some noise and reduce the likelihood of over-fitting occurring. When I compute cross-validation, the model in the recall part of the process; the model is not in the training phase. I use every one of the features that the network has to offer. The model loss curve is shown in Fig. 4.

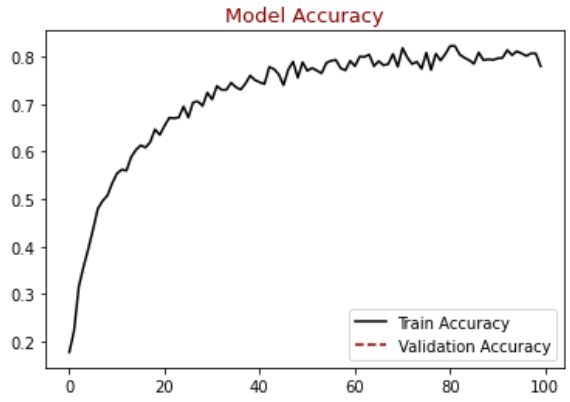

Figure 3. Model accuracy.

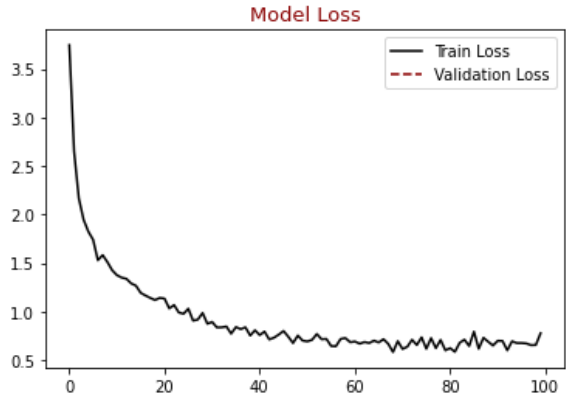

Figure 4. Model loss.

In addition, I used the model to make predictions for a few images taken from the whole dataset that had labels in order to validate the validity of the model. Although there are rare instances in which our model forecasts the incorrect category of activity, on the whole, the performance of our model is satisfactory. When I repeat the code, our model pulls various pictures from the dataset to test and verify our train model. The confusion matrix is an accomplishment indicator for machine learning that categorizes problems that have several possible solutions. The following four distinct sequences of expected and original findings are shown in Fig. 5.

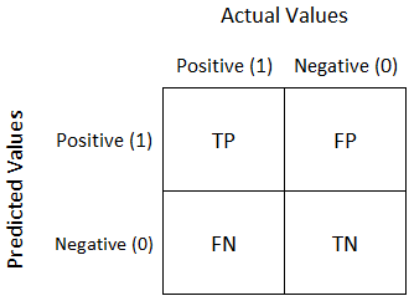

Figure 5. Confusion matrix values.

Confusion matrix is great for analyzing things like recall, precision, specificity, accuracy, and, most crucially, AUC-ROC curves.

True Positive (TP): The predicted value of our model is the same as the actual value; the actual positive value is predicted by our model as positive.

True Negative (TN): The predicted value of our model is the same as the actual value; the actual negative value is predicted by our model as negative.

False Positive (FP): Our model falsely predicted the value; the actual value is negative but our model predicts it as positive.

False Negative (FN): Our model falsely predicted the value; the actual value is positive, but our model predicts it as negative.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

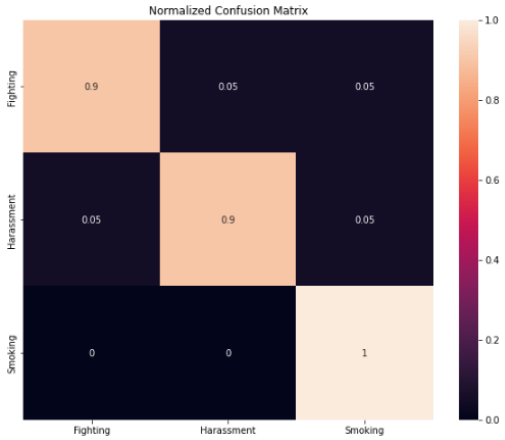

Figure 6. Normalized confusion matrix.

Fig. 6 shows that our model predicts fighting 90% correct, harassment 90%, and smoking 100%. Also, our model predicts some wrong classes, like actual is a fight, but 0.05% of our model predicts harassment, same as actual is smoking, but the model predicts 0.5% harassment.

4.1. Class Activation Heat-Map for Image Classification

Class activation mapping is a mechanism for producing heat-maps of images that demonstrate which parts of a neural network for image classification are of significant value. Score-CAM and Grad-CAM are two of the many versions of the approach (Gradient Weighted Class Activation Mapping). The heat-maps produced by CAM are a visualization that can be understood as indicating exactly in the photo the neural network is (metaphorically) seeking to make its choice. It is your obligation as a deep learning expert to ensure that your model is working successfully. Debugging your model and visually validating that it is "looking" and "activating" at the right places in an image is one approach to do this. Grad-CAM generates a heat-map representation for each class label. This heat-map can be used to graphically confirm where the CNN is focusing in the picture.

4.2. Grad-CAM Class Activation Visualization

Grad-CAM class activation visualization is depicted in Fig. 7.

Figure 7. Normalized confusion matrix.

5. Conclusion

The number of surveillance cameras positioned to monitor private and public places and regions have increased significantly. The process of processing a video, capturing data, and analyzing the data to obtain domain-specific information is called video analysis. Nowadays many CCTV cameras are used to record video but they do not know what happen actually. Some of the systems are used to detect human activities like walking, jogging, setting, etc. This project aims to create a surveillance security system that detects human Deviant and Criminal activity to enhance the security of public surveillance. However, algorithms that analyze captured images and detect deviant circumstances. We have used deep learning conventional neural network techniques to train our model with different images of different activities of humans. Our model predicts that 93.33% correct the deviant classes in testing and validating procedures. This method focuses on identifying deviant actions by training a model for different types of deviant and criminal conduct using photographs. Using a different source like CCTV cameras and Google, to collect data to train our model. We have used a different style of the human image and a different background of the image. To reduce noise and improve the accuracy of a model that uses rotation, flipping the image, zoom, image data generator, and image rescaling to get better results from natural changes like lighting and shadows. Our model trains on real-time deviant activity such as smoking, harassment, and fighting. In the future, we have deployed this model on the CCTV operating system to detect real-time Deviant and Criminal activities. Where the input video will be captured by one fixed camera in different formats with different frames per second, and the input video will be captured indoors or outdoors. Our purpose model is very helpful for security purposes in private or government sectors like airports, stations, banks, and any place.

References

[1] "Data generated by new surveillance cameras to increase exponentially in the coming years | Security Info Watch." Accessed: Feb. 02, 2023. [Online]. Available: https://www.securityinfowatch.com/video-surveillance/news/12160483/data-generated-by-new-surveillance-cameras-to-increase-exponentially-in-the-coming-years

[2] S. O. Ba and J.-M. Odobez, "Recognizing Visual Focus of Attention from Head Pose in Natural Meetings," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 39, no. 1, pp. 16–33, Feb. 2009, doi: 10.1109/TSMCB.2008.927274.

[3] "Video Surveillance and VSaaS Market." Accessed: Feb. 01, 2023. [Online]. Available: https://www.transparencymarketresearch.com/video-surveillance-vsaas-market.html

[4] N. M. Nayak, R. J. Sethi, B. Song, and A. K. Roy-Chowdhury, "Modeling and Recognition of Complex Human Activities," in *Visual Analysis of Humans: Looking at People*, T. B. Moeslund, A. Hilton, V. Krüger, and L. Sigal, Eds., London: Springer, 2011, pp. 289–309. doi: 10.1007/978-0-85729-997-0_15.

[5] S. Rankin, N. Cohen, K. Maclennan-Brown, and K. Sage, "CCTV Operator Performance Benchmarking," in *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*, Oct. 2012, pp. 325–330. doi: 10.1109/CCST.2012.6393580.

[6] T. S. Bora and M. D. Rokade, "Methodology for Human Suspicious Activity Detection," vol. 6, no. 5, p. 4, 2021.

[7] M. Ramzan, A. Abid, and S. Mahmood Awan, "Automatic Unusual Activities Recognition Using Deep Learning in Academia," *Comput. Mater. Contin.*, vol. 70, no. 1, pp. 1829–1844, 2022, doi: 10.32604/cmc.2022.017522.

[8] R. Nawaratne, D. Alahakoon, D. De Silva, and X. Yu, "Spatiotemporal Anomaly Detection Using Deep Learning for Real-Time Video Surveillance," *IEEE Trans. Ind. Inform.*, vol. 16, no. 1, pp. 393–402, Jan. 2020, doi: 10.1109/TII.2019.2938527.

[9] H. Liang and W. Li, "Personalized Recommendation Algorithm for University Civics Courses with Multiple User Interests," *Math. Probl. Eng.*, vol. 2022, pp. 1–10, Mar. 2022, doi: 10.1155/2022/5207167.

[10] F. García-Sánchez, R. Colomo-Palacios, and R. Valencia-García, "A social-semantic recommender system for advertisements," *Inf. Process. Manag.*, vol. 57, no. 2, p. 102153, Mar. 2020, doi: 10.1016/j.ipm.2019.102153.
[11] Z. Gharibshah, X. Zhu, A. Hainline, and M. Conway, "Deep Learning for User Interest and Response Prediction in Online Display Advertising," *Data Sci. Eng.*, vol. 5, no. 1, pp. 12–26, Mar. 2020, doi: 10.1007/s41019-019-00115-y.
[12] J. Shokeen and C. Rana, "Social recommender systems: techniques, domains, metrics, datasets and future scope," *J. Intell. Inf. Syst.*, vol. 54, no. 3, pp. 633–667, Jun. 2020, doi: 10.1007/s10844-019-00578-5.