ISSN 2278-3091
Volume 11, No.2, March - April 2022
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse051122022.pdf
https://doi.org/10.30534/ijatcse/2022/051122022

A Review of Computer Vision Techniques for Video Violence Detection and intelligent video surveillance systems

Imane Rahil[1], Walid Bouarifi[2], Mustapha Oujaoura[3]
[1] Mathematical Team and Information Processing, National School of Applied Sciences SAFI, Cadi Ayyad University, Morocco, imane.rahil@uca.ac.ma
[2] Mathematical Team and Information Processing, National School of Applied Sciences SAFI, Cadi Ayyad University, Morocco, w.bouarifi@uca.ac.ma
[3] Mathematical Team and Information Processing, National School of Applied Sciences SAFI, Cadi Ayyad University, Morocco, oujaoura@gmail.com

Received Date : February 10, 2022 Accepted Date : March 12, 2022 Published Date : April 06, 2022

ABSTRACT

With the advancement of technology in recent years, intelligent video surveillance systems are increasingly being used a large number of cameras, which have been installed in public spaces; such as market places, shopping malls, hospitals, banks, streets, education institutions, city administrative offices, and smart cities. This is to make public lives and assets safer and to keep the risk at a tolerable level. Such systems are being continuously developed due to the advancements in the computer vision techniques. This paper presents an overview of the most used computer vision techniques contemporary issues of the intelligent video surveillance systems for public spaces described in the world's scientific publications. In this paper, the methods of detection are divided into four main categories that are based on classification techniques used: violence detection using traditional machine learning, using Deep Learning, using Convolutional Neural Network (CNN) and using Support Vector Machine (SVM). The feature extraction techniques and object detection techniques of each single method are also presented. Essential basic concepts for computer vision and violence detection are also summarized.

Key words : Computer vision, violence detection, machine learning, deep learning, CNN; SVM ..

1.INTRODUCTION

Recently, using computer vision techniques, the challenge of human violence action tracking at a distance has been tractable [1]. Because of its several practical applications, violence detection has been gaining traction as a research topic [2], [3]. since, unfortunately, violent scenes in movies and media have become common. In addition, because young people have easy access to these media content, violence and crime have increased at an alarming level due to factors primarily related to the influence of the media on people's behavior [4]. An immediate intervention is required to study, analyze, and recognize violent activities in public areas (banks, markets, streets, buses, hospitals, institutions, etc...) [5]. Violent activities include a wide range of behaviors such as vandalism, explosions, and fighting [6], [7].

Surveillance camera footage are processed using computer vision techniques for violence detection [8]. Surveillance and monitoring are observing people's behavior to assess whether their actions are suspicious or legitimate [9]. It is a hard task to regularly detect suspicious activity or to identify such activity across massive data sets containing video recordings [10].

Various approaches for recognizing human actions in real life have been established. These techniques enable the detection of abnormal behaviors in surveillance cameras collected in public places [11].

The recognition of violence from surveillance footage is indeed a category of activity detection [12]. To detect violent situations and other dangerous behaviors in videos, several approaches and techniques can be applied [13]. Particular techniques are suggested in these systems, each of which operates with a different set of input criteria, which are essentially multiple features or elements of the video, such as acceleration, flow, time, appearance [14], [15]. The first phase in recognizing violent activity is to segment and frame a video. Second, it finds an item in video sequence. Next, according to the approach used, extract the video's attributes. Finally, it examines the frames for suspicious behavior [16].

The procedure differs depending on the level of detecting technique used. Figure 1 depicts the fundamental processes of violence detection systems. Many scientists presented many approaches to improve the detection process' efficiency, precision, and quality [17], [18].

Using an extensive literature review, this research investigates and analyzes various methods for recognizing violence from surveillance cameras using computer vision. The basic purpose of this work is to provide an in-depth, detailed evaluation of the techniques for detecting violence. Numerous methods of recognizing violence and unusual behavior have been presented throughout the previous decade.

To carry out a comprehensive research work, we used basic search keywords to collect the most related studies on the detection of violent behavior from library resources. Based on their performance, we want to compare and examine computer vision techniques.

Researchers have increasingly been attracted to the area of violence detection research. To learn the regular patterns in the training footage for anomaly detection, reference [19] suggests a convolutional spatio-temporal autoencoder. reference [20] describes another approach to this model, in which a convolutional long short-term memory (CLSTM) is applied to build the model for activity recognition. For violence detection and recognition, reference [21] employed an extension of improved Fisher vectors (IFVs), which enable videos to be represented using both local attributes and their spatiotemporal locations.

The following is the content of the systemic review:

1) highlights the most recent and promising feature computer vision techniques utilized in anomaly and violence detection tasks,
2) reviews current studies in anomaly detections using computer vision techniques and a few methods that have shown to be promising for our goal.

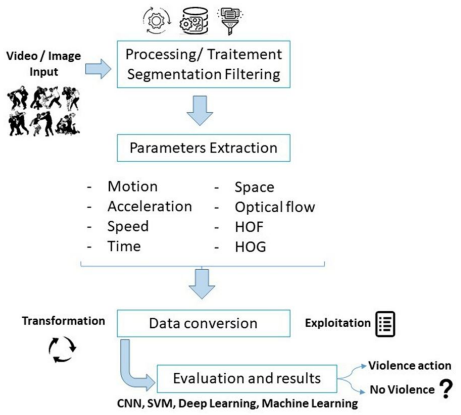

Figure 1: The fundamental processes of violence detection systems

BASIC FEATURES

This section discusses about the basic features related to violence, computer vision, video features used in this state of art.

Many academics are drawn to the science of computer vision because of the large range of applications combining image and video processing. The detection of objects in images and videos, as well as the detection of the object's activity, are all part of image and video analysis. The overall process of activity recognition begins with the shooting of an image or editing of a video.

Table 1: Basic features related to computer vision and violence detection

Features	Description
Computer vision	It includes a set of techniques that empower the computer to "see" and extract information from what it sees.
Motion	The change in the position of the body in space in a video.
Speed	It is the change of its position over time.
Acceleration	It is the speed variation of the image over time
Optical flow	It is a technique used to describe image motion.
HOF; Histogram of optical flow	It is based on optical flow information to describe the normal patterns on the scene.
HOG; Histogram of oriented gradients	It is a feature descriptor used for the purpose of object detection.
Violence	It is an abnormal act in which we can use of, physical or mental, force or power to coerce, dominate, kill, destroy or damage.

2.COMPUTER VISION TECHNIQUES

Computer vision is a discipline of science that deals with tasks related to the image and video analysis. We can assume that in each of them it is necessary to answer the question of what is shown in the picture [22]. Despite the trivial appearance of the question, it is not that easy to answer. There are several computer vision techniques, we can mention in this review among others as shown in figure 2:

A. Image Classification

Image Classification is a fundamental task that attempts to comprehend an entire image as a whole. The goal is to classify the image by assigning it to a specific label. Typically, Image Classification refers to images in which only one object appears and is analyzed [23]. In contrast, object detection involves both classification and localization tasks, and is used to analyze more realistic cases in which multiple objects may exist in an image

B. Object Detection

Object detection is most likely the most profound and complex aspect of computer vision in view of the myriad of practical cases. Object detection refers to the ability of a computer and software systems to find objects in an image / scene and to identify each object. This technique is widely used for face detection, vehicle detection, pedestrian counting, web images, security systems and driverless cars [24]. There are many ways to use object detection technology, as well as many areas to explore. As with any computer technology, a wide range of amazing and creative applications of object detection technology will certainly come from programmers and software developers [25].

C. Object Tracking

Tracking (or tracking objects) is one of the important mechanisms of modern video surveillance systems, which allows you to count objects (people, cars, etc.) and keep track of their movements.

Tracking methods have appeared quite a long time ago [26]. This is, for example, the standard Lucas-Kenedy algorithm for tracking points, which matches points in different frames with each other.

The quality of tracking work rests on the quality of object identification in the video. These were originally HOG descriptors. Then higher quality features - SURF, SIFT, FREAK, etc [27].

Modern positioning systems collect data on the state of the environment and allow you to instantly receive information

about the level of lighting, temperature, pressure, humidity, radiation and the concentration of various substances in the air.

One can use equipment for monitoring vital individual indicators: pressure, heart rate, body temperature. You can add many attributes to positioning devices, organize devices into groups and define zone roles, send push notifications to selected users, and organize local voice communications. Effectively combining a real-time positioning system with video surveillance and radio communication [28]. For example, you can contact a worker whose tag sent an alarm to the console when entering a high-risk area.

D. Semantic Segmentation

Semantic segmentation is the process of breaking up an image into its component parts (segments) and simultaneously classifying these parts. The splitting is done in such a way that when you combine all the parts together, you get the original image.

Each segment of the image consists of pixels; the pixels of one segment have the same class label, which indicates that this segment belongs to the class of segments united by some common characteristics, for example, texture, color, brightness, the presence of repeating elementary objects, etc. Examples of segment classes: road, car, sky, pedestrian, building, animal, bicycle, etc. An example of how semantic segmentation works [29].

Accordingly, the task of semantic segmentation is to find areas (segments) on the image, as well as their classification according to a predetermined set of classes. There is no universal method for solving this problem, so most often the choice of the method is based on the subject area in which this task is posed. However, in the last decade, research and frequency of use (especially for road segmentation) a method based on pixel-by-pixel segmentation came out. Per-pixel segmentation consists in a separate classification of each pixel of the image, and on the basis of this, the restoration of the segments is already taking place [30]. The classification of each pixel most often takes into account the surrounding pixels, that is, not independently. The segmentation method depends on the specific set of methods that will be discussed in the following chapters.

E. Instance Segmentation

Instance segmentation is a computer vision technique that is used to detect objects using the image segmentation technique. It identifies every instance of objects present in images or videos at the pixel level. In image segmentation, visual input is divided into segments to represent an object or part of objects by forming a set of pixels. Instance segmentation identifies each instance of every object represented in the image, instead of categorizing each pixel as in semantic segmentation [31].

Instance segmentation is closely related to two important tasks in computer vision, namely semantic segmentation and object detection. The purpose of semantic segmentation is to label each pixel according to its object class. However, this method does not detect differences between two different object instances of the same class [32]. For example, if there are two people in the image, semantic segmentation will assign the same label to pixels belonging to either of those two faces.

Segmentation of object instances is a relatively new area of computer vision. The existing work on segmentation of an object instance can be classified into two categories: instance segmentation at the detection level and instance segmentation at the image level. Detection-level instance segmentation techniques generally consist of two stages: object detection and semantic segmentation. This method takes into account all generated instances in the image and omits overlaps between different instances. In other words, a pixel in an image can belong to the segmentation masks of two different object instances. Whereas segmentation of an instance at the image level aims to assign each pixel to a maximum of one instance of an object in the image. Since this method must assign each pixel to a unique object instance, this is more complex than segmentation of detection level instances [33].

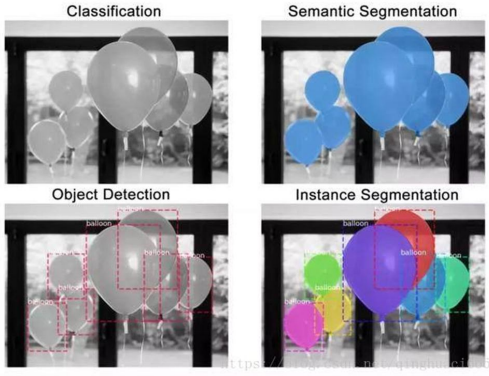

Figure 2: Various Computer Vision Techniques

3.VIOLENCE DETECTION TECHNIQUES

Suspicious actions or behaviors in daily life are considered to be acts of violence. In the discipline of action detection, computer vision recognition of such behaviors in surveillance videos has become a high priority.

As the prevalence of crime has increased, several studies have suggested various techniques and methods for detecting violent or unusual incidents. Various methods for detecting violence that have been developed in recent years are discussed.

A. Machine learning

Nowadays, human behavior is increasingly being recorded using digital cameras [34]. Following this increase of video data, there is a growing need for the development of intelligent video analysis systems, capable of providing value in several real-world domains, including video surveillance, human-robot interactions, entertainment and health applications and violent action detection, which is potentially useful to assist security personnel or to perform emergency calls [35].

In Machine Learning, a computer program is assigned to perform some tasks and it is said that the machine has learnt from its experience if its measurable performance in these tasks improves as it gains more and more experience in executing these tasks [36]. Therefore, the machine takes decisions and does predictions / forecasting based on data [37]. Take the example of computer program that learns to detect / predict violent action from the videos of people from public area [38]. It will improve in performance as it gathers more experience by analyzing images and videos of wider

population. Its performance will be measured by the count of correct detection of violent activities.

An interesting trend is emerging in which ready-to-use machine learning algorithms for speech recognition, language translation, text classifications, and a variety of other tasks are being developed. They are now available as web-based services on cloud computing platforms, expanding the community of developers who can build high-level machine learning systems [39].

Machine learning algorithms are divided into two categories in general. The first is supervised learning, in which the aim is to predict a certain output variable that is linked to each input item. Unsupervised learning is the second main type of machine learning methods, in which the incoming data does not have any labels to go with it. In a method known as clustering, such issues are tackled by identifying some relevant structure in the incoming data. We can tackle issues with little or no knowledge of the outcome if we use unsupervised learning [40].

The first step in using machine learning to solve a problem is to figure out how to translate the learning issue into an algorithm that the computer can understand. The second stage is to choose an assessment technique that assigns a quality or accuracy score to a machine learning algorithm's prediction, usually a classifier. A good classifier will have a high level of accuracy, which means it will make a prediction that is close to the accurate, true label a large proportion of the time. The application of machine learning to address a problem is the third phase [41].

Table 2: Different Machine Learning Algorithms

Machine Learning Algorithms			
Linear Regression	**Non-Linear Regression**	**Linear Classification**	**Non-Linear Classification**
Ordinary Least Squares Regression	Multivariate Adaptive Regression Spines (MARS)	Logistic Regression	Mixture Discriminant Analysis (MDA)
Stepwise Regression	Suport Vector Machine (SVM)	Linear Discriminant Analysis (LDA)	Quadratic Discriminant Analysis (QDA)
Principal Component Regression	K-Nearest Neighbor (kNN)	Partial Least Squares Discriminant Analysis	Regularized Discriminant Analysis (RDA)
Partial Least Squares Regression	Neural Network		Neural Network, Flexible Discriminant Analysis (FDA)
Ridge Regression	Classification and Regression Tress (CART)		Suport Vector Machine (SVM)
Least Absolute Shrinkage Selection Opeartor (LASSO)	Conditional Decision Trees		K-Nearest Neighbor (kNN)
ElasticNet	Modal Trees		Naive Bayes
	Rule systems		Classification and Regression Tress (CART)
	Bagging CART		C4.5
	Random Forest		PART
	Gradient Boosted Machines (GBM)		Bagging CART
	Cubist		Random Forest
			Gradient Boosted Machines (GBM)
			Boosted C5.0

B. Deep learning

The notion of Deep Learning (DL) initially introduced in 2006 as a new discipline of machine learning research [42]. It was initially known as hierarchical learning, and it generally encompassed a wide range of pattern recognition study domains. Deep learning focuses on nonlinear processing in numerous layers or stages, as well as supervised or unsupervised learning. The term "nonlinear processing in several layers" refers to a method in which the current layer uses the preceding layer's result as an input. Layers are organized into hierarchies to arrange the relevance of the data that will be regarded beneficial or not. On the other hand, the class target label is linked to supervised and unsupervised learning; its presence indicates a supervised system, while its absence indicates an unsupervised system [42].

Deep learning includes hierarchical approaches and abstract layer analysis. It may, however, be put to a variety of practical uses. For instance, gray scale image coloring from a photo used to be done manually by users who had to pick each color based on their own perception in digital image processing [43]Coloring may be done automatically by a computer using a deep learning system. Similarly, utilizing Recurrent Neural Networks (RNN) as part of the deep learning techniques, sound may be added to a silent drumming video [44]. Deep learning is a way for improving outcomes and reducing processing times in a variety of computer operations. Deep learning algorithms have been used for picture caption production and handwriting generation in the field of natural language processing. In pure digital image processing, medicine, and biometrics, the following applications are classified [45].

In 2006, the greed algorithm and hierarchy were coupled to create a handwritten digits processing program. Deep learning has recently been used as the primary method for digital image processing in studies. Convolutional Neural Networks (CNN) for Iris Recognition, for example, can be more effective than traditional iris sensors. The efficacy of CNN

can reach 99.35 % [46]. Identification based on a face image has recently been replaced by automated recognition using age and gender as basic characteristics. Sighthound Inc., for example, put to the test a deep convolutional neural network algorithm that could detect not only age and gender, but also emotions [47]. Furthermore, using a deep multi-task learning architecture, a robust system was constructed to properly detect a person's age and gender from a single photograph [48].

Deep learning is utilized in conjunction with biometric features in terms of security, particularly access control. FaceSentinel facial recognition devices were developed and optimized with the help of DL [49]. In nine months, this firm claims, its gadgets will be able to increase their identification process from one–to–one to one–to–many. Without the introduction of DL, this engine innovation may have taken ten man years. It has sped up the equipment's manufacture and launch [44].

Deep Learning is a type of machine learning that use numerous layers of processing to learn non-linear data representations at high degrees of abstraction [50]. In the disciplines of computer vision and natural language processing, designs such as Deep Neural Networks, Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) have been used to achieve state-of-the-art outcomes in a variety of applications [51].

The capacity to produce features automatically from accessible data is the fundamental benefit of these models, permitting pattern recognition systems to focus less on manually-built heuristics.

Despite the fact that Convolutional Neural Networks (CNNs) have been showing great results in handwritten digit classification and face recognition activities since the late 1990s, CNNs have recently shown excellent performance on more challenging visual classification tasks, most particularly with [52] winning the 2012 ImageNet classification benchmark with a CNN model attaining an error rate of 16.4% compared to second place result of 26.1 %.

Reference [53] attributed this improvement to the availability of much larger training sets, which increased the number of labelled data significantly; more powerful GPU implementations, which enabled the training of very large models; and better model regularization strategies, such as Dropout, which prevented overfitting.

C. CNN (Convolutional neural networks)

An artificial neural network is made up of a series of simple, connected components known as neurons that produce a set of real-valued activations [54]. Neurons are activated by weighted links from the active neurons of the previous layer. Learning is influenced by neuronal forwarding of information. Obtaining weights that enable neural networks to display the desired behavior is the purpose of learning. Depending on the activity and orientation of neurons, several computing steps may be required, with each stage allowing the network to learn by modifying weights, referred to as the training phase [55].

Convolutional neural networks (CNNs) have recently improved in accuracy and performance for a wide range of computer techniques, including behavior detection and security, object tracking and activities recognition, videos summarization, and disaster risk management [56].

The most of computer vision algorithms are constructed using CNN. A convolutional neural network is a deep learning technique that takes an input data and assigns value to various objects in the image, allowing them to be distinguished from one another [57].

CNN does not require as much preprocessing as other approaches. As a result, the most effective learning algorithm for interpreting image information is a CNN [58]. It also performs exceptionally well in image classification, identification, segmentation, and information extraction.

Where there is enough labelled data available for training, CNN has usually proven great effectiveness in traffic sign identification, medical picture segmentation, face detection, and object identification in natural images [59]. Fig. 3 depicts a number of CNN uses. On the basis of the data set and retrieved features, a convolutional neural network is utilized to categorize the violent recognition employing additional convolutional layers [60]. Typical fight detection methods use on field knowledge to create complicated handcrafted attributes from the input [61]. Deep models, on the other hand, can respond immediately and extract features automatically.

CNN collects inputs through layers and transmits information through nodes [62]. However, unlike a simple neural network, its layers are highly specialized and can absorb large amounts of input (image and video). There are four types of layers in this network: convolution, ReLu, pooling, and fully connected (FC) [63]. A filter or kernel is slid across the volume of the convolution layer, and the convolution process is conducted to create an activation map. The pooling layer receives the generated volume, which is required to collect the defining features from the preceding layers. A maxed-pooling or averaged pooling layer can be used as the pooling layer. After that, the matrix is flattened to produce a one-dimensional column vector, which is then sent to the FC layer. The FC layer serves as a classification layer. There are many variants of CNNs but the basic architecture of CNN remains the same. The AlexNet, whose design is shown in Fig. 4, is one of the most important CNNs.

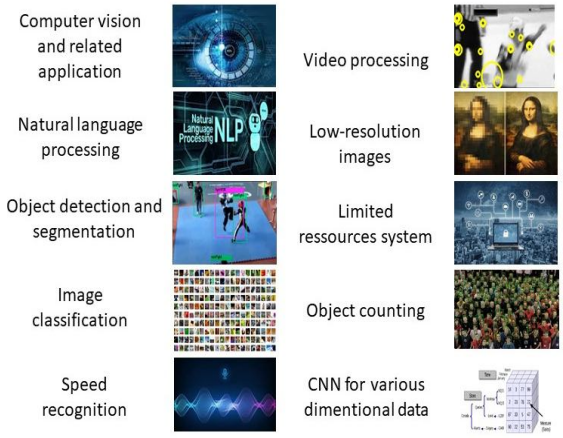

Figure 3: Applications of CNN.

Figure 4: The AlexNet architecture for image recognition

Other variations have been suggested. CNN architectures, as well as other deep learning architectures, have been introduced to improve the classification of violent acts [64]. Deep learning architectures, which are similarly based on neural networks, use further convolution layers to classify violent acts based on the dataset and retrieved attributes [65]. When we deconstruct a video, we find that it is made up of multiple frames of images. The illusion of constantly moving images is what video is. We can identify the scenario of a video as well if we can classify the image in the frame as either violent or non-violent.

Reference [66] proposed a 3D convolutional neural network for action recognition, stacking multiple contiguous frames of video and using as input for the network, capturing the motion information. This method achieved similar results to those using dense trajectories, but computing over frames with a 4-times lower resolution.

In 2015, reference [67] presented GoogLeNet, a deep CNN in which all filters in the architecture are learned and the layers repeated several times, leading to a 22-layer deep network that obtained excellet performance in the 2014 ImageNet classification benchmark. This challenge involved the classification of images into one of 1000 possible categories, and GoogLeNet achieved an error rate of 6.67%, compared to the second place result of 7.32%.

D. SVM (Support video machine)

SVM is a learning technique that accepts the data as support vectors and constructs a hyperplane to classify them (Fig. 5) [68]. It is mostly used for classification issues. In an SVM (Fig. 6), three types of kernels are used: linear, polynomial, and radial basis functions (RBF). When the data are linearly separable, the linear kernel is effective; however, the polynomial kernel is better for data that can be separated by a polynomial degree curve. The RBF kernel generates the hyperplane using the squared Euclidean distance between two vectors. Reference [69] used the ViF descriptor to characterize the shift in flow-vector magnitudes and a linear SVM to detect violence.

Because it is resilient and takes quantitative characteristics into consideration, SVM is a frequent strategy in computer vision [70]. It is used to do binary classification. The kernel is the basis of SVM. Kernel is a function that converts data into a high-dimensional space that may be used to solve a problem. A major drawback of SVM is the lack of transparency in the results.

A Support Vector Machine (SVM) is a supervised learning technique that may be used to solve a variety of classification and regression issues, including medical signal processing, natural language processing, speech identification, and visual recognition of violence [71]. The support vectors are a subset of learning observations that determine the separation hyperplane's position. For binary classification tasks, the conventional SVM algorithm is used.

Typically, multi-class issues are broken down into a sequence of binary problems.

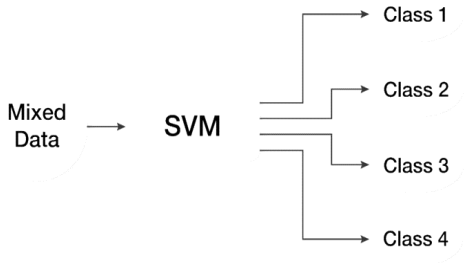

Figure 5: SVM classifier

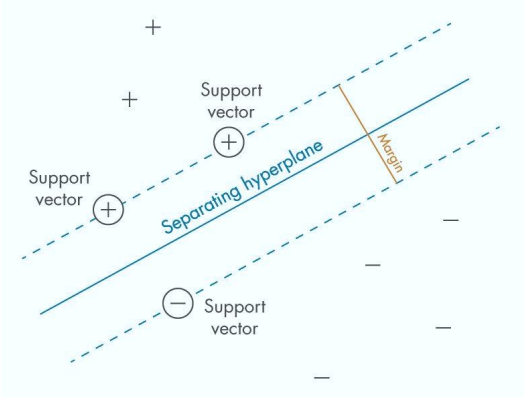

Figure 6: SVM operating principle

4.CONCLUSIONS

Recently, intelligent video surveillance has become widespread in various fields of human activity. If earlier neural networks were used mainly to ensure safety in transport, strategically important state facilities and public places, today they are being actively implemented in different areas to detect some abnormal activities. The powerful development of analytical algorithms has made it possible to significantly expand the range of real scenarios for the use of equipment. Indeed, many researchers proposed different techniques for detection of such activities. The main goal of this review is to explore the state of the art research in the computer vision techniques and violence detection system. Furthermore, a comparative and detailed study of the different techniques of violence and abnormal activities detection has been developed in this present work Moreover, datasets and video features that used in all techniques, which are important in the recognition process are assigned and detailed in a comprehensive table.

REFERENCES

[1] C. Okinda, I. Nyalala, T. Korohou, C. Okinda, J. Wang, T. Achieng, ..., and M. Shen, "A review on computer vision systems in monitoring of poultry: A welfare perspective", *Artificial Intelligence in Agriculture* 2020.

[2] H. Yao and X. Hu, "A survey of video violence detection", *Cyber-Physical Systems*, pp. 1-24, 2021.

[3] M. Mudgal, D. Punj, and A. Pillai, "Suspicious action detection in intelligent surveillance system using action attribute modelling", *Journal of Web Engineering*, pp. 129-146, 2021.
[4] J. He and H. Zheng, "Prediction of crime rate in urban neighborhoods based on machine learning". *Engineering Applications of Artificial Intelligence*, vol. *106*, 104460, 2021.
[5] J. Ren, F. Xia, Y. Liu, and I. Lee, "Deep Video Anomaly Detection: Opportunities and Challenges". *arXiv preprint arXiv:2110.05086*, 2021.
[6] J. Mateo, "4. Street Gangs of Honduras. In *Maras"*, University of Texas Press, pp. 87-104, 2021.
[7] P. M. Pannirselvam, M. K. Geetha, and G. Kumaravelan, "A Comprehensive Study on Automated Anomaly Detection Techniques in Video Surveillance". *Annals of the Romanian Society for Cell Biology*, pp. 4027-4037, 2021.
[8] F. U. M. Ullah, M. S. Obaidat, K. Muhammad, A. Ullah, S. W. Baik, F. Cuzzolin, ...and V. H. C. de Albuquerque, "An intelligent system for complex violence pattern analysis and detection", *International Journal of Intelligent Systems*, 2021.
[9] K. Rezaee, S. M. Rezakhani, M. R. Khosravi, and M. K. Moghimi, "A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance". *Personal and Ubiquitous Computing*, pp. 1-17, 2021.
[10] L. Von Ziegler, O. Sturman, and J. Bohacek, "Big behavior: challenges and opportunities in a new era of deep behavior profiling". *Neuropsychopharmacology*, vol. *46*(1), pp. 33-44, 2021.
[11] J. M. Ackerson, R. Dave, and N. Seliya, "Applications of recurrent neural network for biometric authentication & anomaly detection". *Information*, vol. *12*(7), 272, 2021.
[12] A. Mehmood, "Abnormal Behavior Detection in Uncrowded Videos with Two-Stream 3D Convolutional Neural Networks", *Applied Sciences*, vol. *11*(8), pp. 3523, 2021.
[13] A. Nassauer and N. M. Legewie, "Video data analysis: A methodological frame for a novel research trend". *Sociological methods & research*, vol. *50*(1), pp. 135-174.
[14] A. Srivastava, T. Badal, A. Garg, A. Vidyarthi, and R. Singh, "Recognizing human violent action using drone surveillance within real-time proximity", *Journal of Real-Time Image Processing*, vol. *18*(5), pp. 1851-1863, 2021.
[15] S. Coşar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares, and F. Brémond, "Toward abnormal trajectory and event detection in video surveillance", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. *27*(3), pp. 683-695, 2016.
[16] A. B. Mabrouk and E. Zagrouba, "Abnormal behavior recognition for intelligent video surveillance systems: A review", *Expert Systems with Applications*, vol. *91*, pp. 480-491, 2018.
[17] M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection". *IEEE Access*, vol. *6*, pp. 52843-52856, 2018.
[18] N. Khan, A. Ullah, I. U. Haq, V. G. Menon, and S. W. Baik, "SD-Net: Understanding overcrowded scenes in real-time via an efficient dilated convolutional neural network", *Journal of Real-Time Image Processing*, vol. *18*(5), pp. 1729-1743, 2021.
[19] Y. S. Chong, and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder". In International symposium on neural networks, Springer, Cham, June, 2017, pp. 189-196.
[20] S. Sudhakaran, and O. Lanz, August, "Learning to detect violent videos using convolutional long short-term memory", In 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2017.
[21] P. Bilinski and F. Bremond, "Human violence recognition and detection in surveillance videos", 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Colorado Springs, CO, pp. 30-36, 2016.
[22] C. Christensen and S. Corneliussen, "Tracking of Articulated Objects Using Model-Based Computer Vision", Technical report, Laboratory of Image Analysis, Aalborg University, Denmark, June. 1997.
[23] M. Tripathi, "Analysis of Convolutional Neural Network based Image Classification Techniques", Journal of Innovative Image Processing (JIIP), vol. 03(02), pp. 100-117, June. 2021.
[24] S. Bell, C. Lawrence Zitnick, K. Bala, & R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2874–2883.
[25] J. Dai, Y. Li, K. He, & J. Sun, "R-fcn: Object detection via region-based fully convolutional networks". In Advances in neural information processing systems, pp. 379–387, 2016.
[26] W. Luo, J. Xing, A. Milan, X. Zhang W. Liu, and T. K. Kim. "Multiple object tracking: A literature review". Artificial Intelligence vol. 293, pp. 1-32, 2021.
[27] J. Hong Yoon, C.-R. Lee, M.H. Yang, and K.-J. Yoon. "Online multi-object tracking via structural constraint event aggregation", in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Las Vegas, NV, USA, 2016, pp. 1392–1400.
[28] W. Zhang, H. Zhou, S. Sun, Z. Wang, J. Shi, and C.C. Loy, "Robust multi-modality multi-object tracking, in: Proc. IEEE Int. Conf. Comput. Vis., Seoul, Korea, 2019, pp. 2365–2374.
[29] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs", arXiv preprint arXiv:1606.00915, 2016.
[30] Y. Guo, Y. Liu, T. Georgiou and M. S. Lew. "A review of semantic segmentation using deep neural networks. International Journal of Multimedia Information Retrieval", Vol. 7 pp. 87-93, 2018.
[31] L. Yang, Y. Fan, and N. Xu, "Video Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1189–1198
[32] A. Almutairi and M. Almashan, "Instance Segmentation of Newspaper Elements Using Mask R-CNN. In proceedings of the IEEE Conference on Machine Learning and Applications (ICMLA), 2019, pp. 703-707.
[33] J. Cheng, Y.-H. Tsai, S. Wang, & M.-H. Yang, "Segflow: Joint learning for video object segmentation and optical flow". In IEEE International Conference on Computer Vision (ICCV), 2017, pp. 686-695.

[34] N. Almazova, D. Bylieva, V. Lobatyuk, and A. Rubtsova, "Human behavior as the source of data in the education system", In *Proceedings of the International SPBPU Scientific Conference on Innovations in Digital Economy*, October, 2019, pp. 1-7.
[35] D. Nova, A. Ferreira, and P. Cortez, "A machine learning approach to detect violent behaviour from video", In *International Conference on Intelligent Technologies for Interactive Entertainment*, Springer, Cham, November, 2018, pp. 85-94.
[36] S. Ray, (). "A quick review of machine learning algorithms". In the *International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, February, 2019, pp. 35-39.
[37] A. S. Deeks, "Predicting Enemies", *Virginia Law Review*, vol. *104*(8), pp. 1529-1592, 2018.
[38] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos", In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6479-6488.
[39] S. M. Basha and D. S. Rajput, "Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap", *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, Academic Press, pp. 153-164, 2019.
[40] A. Khan, B. Baharudin, L. H. Lee, and K. Khan, " A review of machine learning algorithms for text-documents classification". *Journal of advances in information technology*, vol. *1*(1), pp. 4-20, 2010.
[41] S. Marsland, "*Machine learning: an algorithmic perspective"*. Chapman and Hall/CRC, 2011.
[42] Vargas, R., Mosavi, A., & Ruiz, R. "Deep learning: a review", *Advances in Intelligent Systems and Computing*, 2017.
[43] R. Ilin, T. Watson, and R. Kozma, "Abstraction hierarchy in deep learning neural networks". In the I*nternational Joint Conference on Neural Networks (IJCNN)*, May, 2017, pp. 768-774.
[44] R. Vogl, "Deep learning methods for drum transcription and drum pattern generation", *Johannes Kepler University Linz, Linz*, 2018.
[45] D. L. Yamins, and J. J. DiCarlo, "Using goal-driven deep learning models to understand sensory cortex". *Nature neuroscience*, vol. *19*(3), pp. 356-365, 2016.
[46] T. Lee and M. David, "Hierarchical Bayesian inference in the visual cortex". Journal of the Optical Society of America A, Vol. 20(7), pp. 1434-1448, 2003.
[47] A. Bannister, "Biometrics and AI: how FaceSentinel evolves 13 times faster thanks to deep learning", IFSEC Global, July, 2016.
[48] J. Moor. "The Turing test: the elusive standard of artificial intelligence", Springer Science & Business Media, 2003.
[49] W. Yang, S. Wang, J. Hu, G. Zheng, J. Yang, and C. Valli, "Securing deep learning based edge finger vein biometrics with binary decision diagram", *IEEE Transactions on Industrial Informatics*, vol. *15*(7), pp. 4244-4253, 2019.
[50] H. Patel, A. Thakkar, M. Pandya, and K. Makwana, "Neural network with deep learning architectures", *Journal of Information and Optimization Sciences*, vol. *39*(1), pp. 31-38, 2018.
[51] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks", *Artificial Intelligence Review*, vol. *53*(8), pp. 5455-551, 2020.
[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks", In NIPS, 2012.
[53] M. D. Zeiler, and R. Fergus, "Visualizing and understanding convolutional networks", In *European conference on computer vision*, Springer, Cham, September 2014, pp. 818-833.
[54] T. Minemoto, T. Isokawa, H. Nishimura, and N. Matsui, "Feed forward neural network with random quaternionic neurons", *Signal Processing*, vol. *136*, pp. 59-68, 2017.
[55] I. A. Basheer, and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application", *Journal of microbiological methods*, vol. *43*(1), pp. 3-31, 2000.
[56] S. Mittal, "A survey of FPGA-based accelerators for convolutional neural networks", *Neural computing and applications*, vol. *32*(4), pp. 1109-1139, 2020.
[57] A. Dhillon, and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection", *Progress in Artificial Intelligence*, vol. *9*(2), pp. 85-112, 2020.
[58] R. Zhang, C. Cheng, X. Zhao, and X. Li, "Multiscale mask R-CNN–based lung tumor detection using PET Imaging", *Molecular imaging*, vol. *18*, pp. 1-8, 1536012119863531, Nov. 2019.
[59] Z. Liang, J. Shao, D. Zhang, and L. Gao, "Traffic sign detection and recognition based on pyramidal convolutional networks", *Neural Computing & Applications*, vol. *32*(11), June. 2020.
[60] I. Serrano, O. Deniz, J. L. Espinosa-Aranda, and G. Bueno, "Fight recognition in video using hough forests and 2D convolutional neural network", *IEEE Transactions on Image Processing*, vol. *27*(10), pp. 4787-4797, 2018.
[61] M. Huh, A. Liu, A. Owens, and A. A. Efros, "Fighting fake news: Image splice detection via learned self-consistency", In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 101-117.
[62] A. Rehman, S. U. Rehman, M. Khan, M. Alazab, and T. Reddy, "CANintelliIDS: detecting in-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU", *IEEE Transactions on Network Science and Engineering*, *Feb*. 2021, pp. 1-11.
[63] S. H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling", *Journal of medical systems*, vol. *42*(5), pp. 1-11, 2018.
[64] S. A. Sumon, T. Shahria, R. Goni, N. Hasan, A. M. Almarufuzzaman, and R. M. Rahman, "Violent Crowd Flow Detection Using Deep Learning", In the Asian Conference on Intelligent Information and Database Systems (*ACIIDS) (1)*, April. 2019, pp. 613-625.
[65] M. Marsden, K. McGuinness, S. Little, and N. E. O'Connor, "Resnetcrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification", In *14th IEEE international conference on advanced video and signal based surveillance (AVSS)*, Aug. 2017, pp. 1-7.

[66] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition", *IEEE transactions on pattern analysis and machine intelligence*, vol. *35*(1), pp. 221-231, 2012.
[67] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov,... and A. Rabinovich, "Going deeper with convolutions". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
[68] F. F. Chamasemani, and Y. P. Singh, "Multi-class support vector machine (SVM) classifiers--an application in hypothyroid detection and classification", In *Sixth International Conference on Bio-Inspired Computing: Theories and Applications*, Sept. 2011, pp. 351-356.
[69] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June. 2012, pp. 1-6.
[70] A. Tellaeche, G. Pajares, X. P. Burgos-Artizzu, and A. Ribeiro, "A computer vision approach for weeds identification through Support Vector Machines", *Applied Soft Computing*, vol. *11*(1), pp. 908-915, 2011.
[71] S. Sharma, G. Singh, and M. Sharma, "A Comprehensive Review and Analysis of Supervised-Learning and Soft Computing Techniques for Stress Diagnosis in Humans", *Computers in Biology and Medicine*, 104450, vol. 134, pp. 1-19, 2021.