INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 9 ISSUE 4 (2023) PAGES 1 – 9
RECEIVED : 17.06.2023 PUBLISHED : 17.07.2023

July 17,2023

VIOLENCE DETECTION USING VGG19

M.Ramya[1] and N.Karthik [2*]

[1]Department of Computer Science and Engineering, Kingston Engineering College, Vellore 632007, India; ramyamonali2000@gmail.com

[2]Assistant Professor, Department of Computer Science and Engineering, Kingston Engineering College, Vellore 632007, India; karthik.krthk@gmail.com

Abstract:

CCTVs are frequently employed to prevent crimes from happening in the neighborhood. Although CCTVs are used in many both public and private spaces to monitor the environment, crime control has not improved. This is due to the fact that CCTV needs human monitoring, which can result in human-prone mistakes such as missing some crucial crime incidents while simultaneously monitoring a large number of displays with footage from several CCTV cameras. To solve this problem, we developed the Crime Intention Detection System, which recognizes criminal activity in real-time footage and notifies human supervisors to take appropriate action. To inform the supervisors of any close criminal activity. Our system now includes a MAIL sending module that, whenever violations are discovered, emails the concerned person pointed Pre-trained deep learning algorithm VGGNet-19, which identifies a gun or knife in the grip of a gun pointed at another person, is used to execute the suggested method. Also, we contrasted how two various pre-trained models, such as Google Net InceptionV3, performed throughout training. In terms of learning precision, the outcomes with VGG19 are better. This inspired us to use VGG19 to detect criminal intent in pictures and videos with some little fine-tuning in order to more accurately address the problems with current techniques. An experimental analysis of publicly available datasets that include violent films demonstrates the usefulness of the suggested approach.

Keyword: Violence Detection, Crime Detection, VGG19, Intelligent video surveillance

I. INTRODUCTION:

Pre-trained models using Deep Learning are simulations created to help users to learn more algorithms or test out existing frameworks for improved outcomes without specific design. The roles of these layers will be covered in more detail in the following sections. A deep neural network contains five layers, covering the layers for input and output with a Pooling layer, Max-Pooling, and a Fully connected layer. Several people used to favor Deep Understanding of pre-trained concepts because of some limited time, memory, and facilities like CPU, Processors. And when compared to machine learning, it forces us to construct explicitly, these pre-trained algorithms will produce the best and most accurate outputs. To find firearms in surveillance footage, a lot of human interaction is needed, which increases the risk of human error. Human defenders may get fatigued or nod off when watching films at high volume or while someone is keeping two or more pieces of footage, which causes them to miss some infrequently occurring crime motive scenes. To do this, it is essential to create an automatic monitoring system that can quickly identify firearms and thereby decrease the frequency of crime. The goal of the Crime Motive Identification System is to prevent crimes from happening by identifying people carrying weapons at ATMs, banks, and public places. To do this, the system using pre-trained deep learning algorithms such as Google Net and VGGNet-19. The discussion of pre-trained algorithms will be covered in the section that follows. And we demonstrated that the VGGNet-19 model outperforms the Google Net Fully Convolutional model in regard to training efficiency, improved categorization, and decision-making.

II. REVIEW OF THE LITERATURE

The study of human activity in video has a variety of applications, from security and surveillance to entertainment and personal archiving, and is becoming more and more significant. This problem is difficult to solve due to a number of difficulties at different processing levels, including reliability against bugs at low-level processing, view, and rate-invariant depictions at mid-level processing, and conceptual depiction of human behaviors at higher-level processing. We provide a thorough overview of the work done over the last two decades to tackle the problems of representing, recognizing, and understanding human behavior from video and are prepared to consider it [1]. Human behavior analysis has recently started to branch out into other areas. The development of action detection or action classification approaches has attracted the attention of numerous researchers. The study on evaluating people's behaviors varies from other research in that it aims to develop computation frameworks and evaluation procedures for automatically determining the caliber of human acts. Because of the rapidly expanding practical uses of this field of study, including physical rehabilitation, assisted living for the elderly, skill development on self-learning programs, and sports game scoring, it has gained popularity [2-5]. By minimizing a prejudicial cost operation by slope extraction, training is carried out using both negative and positive samples of a certain action class without the requirement for commentary on the correlation between the training video sequence and the state-conditioned monitors. Simultaneously detecting, monitoring, and recognizing actions, trained models carry out identification and localization. Our method develops intuitively comprehensible models that portray activity as a series of receptive field models, in contrast to several past approaches [6-7]. For many applications that monitor videos, automated violence identification from a video is a hot issue. Nevertheless, there hasn't been much progress in creating an algorithm that can accurately detect abuse in surveillance films. In this study, we suggest two significant improvements to the motion Weber local descriptor (WLD), which we recently suggested as a method for identifying violence in moving images [8-9]. It may be possible to support city administration and urban policy via the integration of video data produced by smart cities. This study proposed a BBM algorithm to assist long-term reference architecture for effective surveillance footage coding. Based on the unique attributes of surveillance footage, such as consecutive pictures having very significant correlation and each image can be partitioned into both foreground and background, this study is based on the particular attributes of surveillance footage. To enhance the coding presentation, a rate-distortion improvement for the surveillance source (SRDO) technique is also created [10-11]. The abnormal or unusual method of analysis in a crowded video scene is exceedingly difficult due to various real-world constraints. The study contains a thorough analysis that progresses from object identification to activity recognition to crowd analysis to violence identification in a crowded context. The vast majority of the articles examined in this survey use deep learning as their foundation. In terms of their structures and algorithms, different deep learning techniques are contrasted. This study's major objective is to apply deep learning algorithms to identify the precise number of participants, involved individuals, and the activities that occurred in a huge crowd under all-weather circumstances [12-15]. Applying localized spatiotemporal description to the inquiry video is a frequent image captioning technique for spotting violence in videos. Finally, using the Bag-of-Words (BoW) paradigm, the low-level explanation is further condensed onto the greater feature. Conventional spatiotemporal descriptions, however, lack sufficient discrimination. Furthermore, the BoW model loosely associates each feature representation with a single visual word, resulting in quantized inaccuracy [16]. Aggression detection is one aspect of the larger issue of activity recognition. In recent years, automated human behavior detection in real videos has grown in significance for applications including surveillance footage, HCI, and content-based video search. The most recent activity recognition techniques can be loosely divided into local, interest-point-based, and worldwide, frame-based techniques. With local approaches, human behavior in a video is represented by spatiotemporal feature points that are found [17-19]. The detection of anomalies and abnormal behavior can be done in a number of ways. The concept of abnormalities, even so, is not one that is concise or clearly defined. We do not concentrate on these methods because abnormality identification is a specific study field with different limits and assumptions [20].

III. PROPOSED METHODOLOGY:

Our system's goal is to create a smart video surveillance system that can recognize violence in a specific video frame. The smart surveillance system trains using previously learned attributes after learning them. It looks for criminal activity in a video sequence, and if any is found in a frame, it will notify the appropriate authorities and record the frame in a local system. A video sequence will serve as the system's input, and its Boolean output is going to be a violent or an inoffensive frame. The device is designed to assist the user by evaluating whether or not the crime happens in a brief video frame. Government entities can respond more quickly with the aid of this method. In order to create a system that can take use of parallel processing for quick processing, we used Tensor Flow GPU packages in our system.

ARCHITECTURE DIAGRAM:

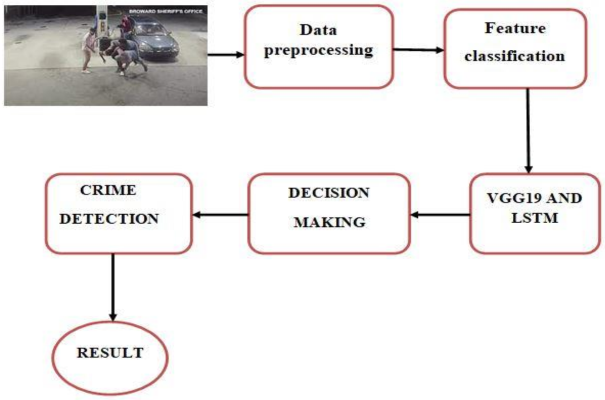


Fig 4.1 Architecture diagram

V. MODULE DESCRIPTION:

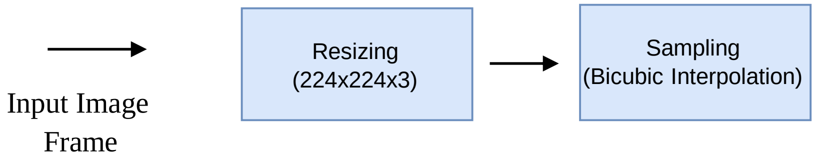
- datasets
- data preparation
- data validation protocol
- decision process

DATASETS:

The designed system is tested using datasets for both photos and videos that are pulled from both Google and YouTube. The data sources belong to the type that includes robberies, murders, and some illegal activities that involve the use of weapons; however, using weapons is strictly forbidden in places like ATMs, banks, and some public areas.

DATA PREPARATION:

The input layer receives input frames and performs preprocessing on them, converting any supplied images (such as those with dimensions of 256*256 to 120*120*3) to regular size (RGB values), by eliminating RGB values again from pixels.



FUTURE EXTRACTION:

A layer that uses Keras can help to derive the characteristics from the Photo Net model's pre-trained values, which were trained using potentially enormous datasets. VGG19, which uses 19 different layer weights, is used to extract features from Image Net. Because the location of the characteristics to be derived in the convolution layer is unknown, this layer draws a filter of 3*3 for all input image pixels and then computes the product of pixels' values as -1*- 1=1 and 1*1=1. Hence, the feature extraction pixel of the test dataset is very near to the values of the qualified pixels. If the comparison of the pixels results in extremely equal to negative numbers or 0, such as -1*1=-1, or 1*-1=-1, these numbers are far from complementing already taught pixels. After each of the pixels inside an input image has been processed through the filtering process, the characteristics are arranged into a 2D array that contains all of the potential features. The computed pixel values are then input to the following layer (like 1,0.3, estimated above as product).

FEATURE MAPPING:

As seen in figure 3.4 below, it is a method used to compare feature extraction to trained characteristics. Every filter that was extracted has a 3*3 form, and after calculations, the dot product numbers were 0, 0.3, 0.4, 0.85, 0.9, and 1.

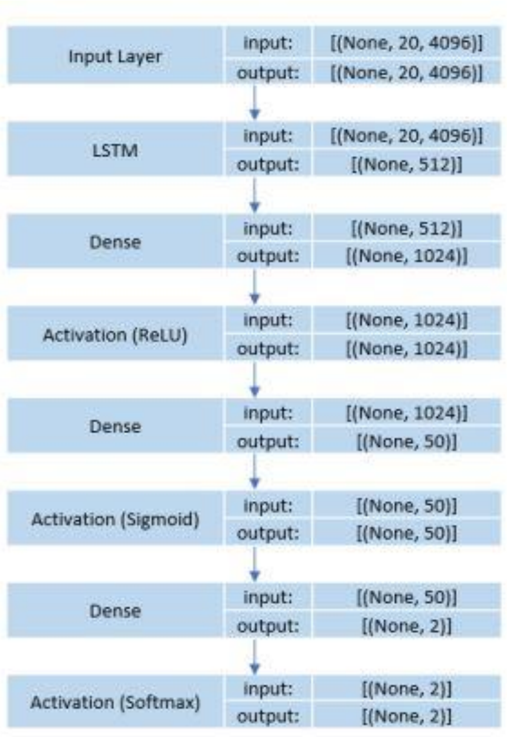
If the values are similar to or equal to 1, then the characteristics are matched to the qualified Image Net dataset, indicating the presence of a feature that is comparable to that of crime.

MAX-POOLING LAYER:

Applying the pooling layer Use the Max-Pooling tool on all photos to decrease the number of pixels and extract the image's most crucial details. To extract small, narrow-depth features from a picture, this technique was repeated.

LSTM :

A little layer performs a crucial role by setting any minus pixel values in images to zero. Any matching positive number of pixels is then given any positive values for the other matched pixels. For feature translations, there are two ReLU stages and a further ReLU layer that after the convolution operation. The training algorithm in the picture below analyzes the pixels from the training model to the pixels from the fresh image; if the function identifies any negative numbers in the pixels, it sets them to 0, or else it leaves them as positive.



MAIL MODULE:

In the event that a person is discovered while committing a crime, the mail module activates and sends an email to the relevant security personnel informing them of the crime's objectives, helping to prevent it from occurring in the initial place.

VI. EXPERIMENTAL RESULTS:

For the Hockey Battle Dataset, the suggested model achieved a 98% accuracy rate. The model's ability to accurately identify violent behaviors or non-violence is used to assess reliability. The formula is used to calculate it.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP -> True Positive,

TN -> True Negative,

FP -> False positive

FN -> False negative

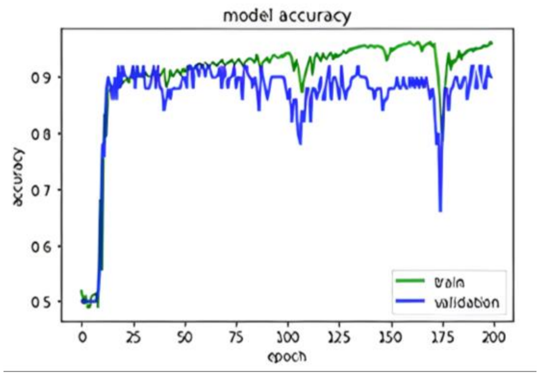


Fig 6.1 Model Accuracy

The figure depicts about training accuracy rate versus validation accuracy rate of the proposed approach.

The figure below displays the method's "Mean Squared Error" lost on the Hockey Combat Dataset. The estimated mean squared error rate with this approach is 0.02.

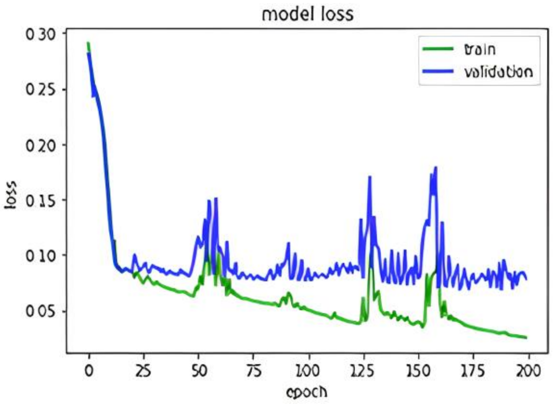


Fig 6.2 Model Loss

The figure depicts about training accuracy rate versus validation accuracy rate of the proposed approach.

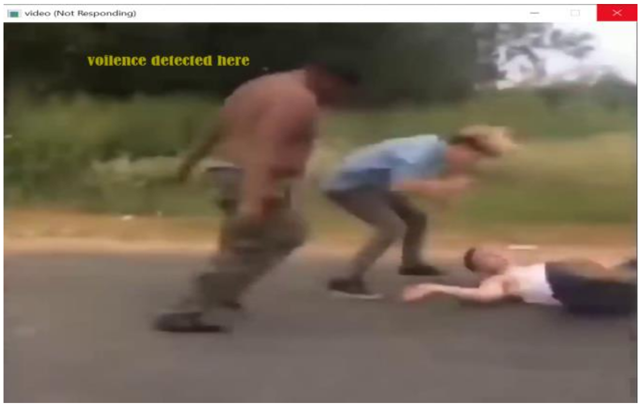
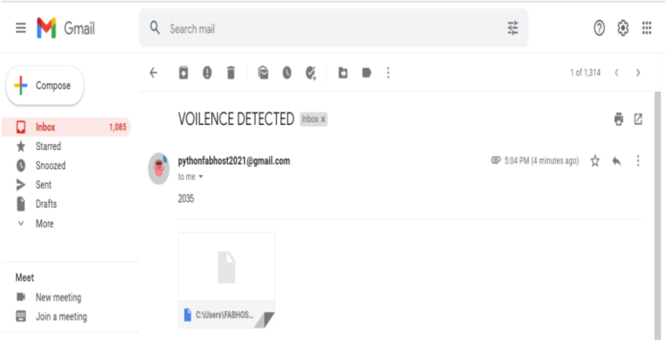


Fig 6.3 Violence Dataset Image

The above mentioned figure 6.3 is one of the violence detection dataset image, when we uploaded or select the image from the application it shows violence detected are not after detected the violence it send to the notification mail to the valid authorities or the admin of the system.



VOILENCE DETECTED

pythonfabhost2021@gmail.com
to me
5:04 PM (4 minutes ago)

2035

C:\Users\FABHOS...

Fig 6.4 Violence detection Notification

CONCLUSION:

The projected crime motive identification system is an automated system that prevents crimes from happening by identifying unusual situations utilizing the pre-trained VGGNET19 architecture in less time than the Google Net Inception V3 model. Our system will generate the criminal intention identification safety message to the authorized mail if a person is found when a crime is committed. Comparing the proposed system to all other criminal Activity detection methods now in use, it produces good results. If a crime is to happen, the added function makes the CCTV send instantly the criminal intention safety messages to emails and the system gets notified. The Developed Crime Intended Identification System features may be embedded in CCTV to identify the crime scenes.

REFERENCES

[1] P. Turga, R. Chellapa, V. S. Subramanian, and O. Udreah, “Machine Detection of human actions: A Review,” IEEE Trans. Circuits System. Video Technology, vol. 18, no. 11, pp. 1473–1488, Nov. 2008.

[2] 1R. Pope, “A Study on vision-based human activities Identification,” Image Vis. Computer., vol. 28, no. 6, pp. 976–990, 2010.Author, F., Author, S., Author, T.: (1999).

[3] S.-R. Ke, H. L.U. Thucc, Y.-J. Lee , J.-N. Huang, J.-H. Yoo, and K.-H. Choi, “A survey on video based human activities Identification,” Computers, vol. 2, no. 2, pp. 88–131, 2013.

[4] I. S. Gracia , O. D. Suarez, G. B. Garcia , and T.-K. Kim, ‘‘Quick violence identification,’’ PLoS ONE, vol. 10, no. 4, Apr. 2015, Art. no. e0120448.

[5] O. Deniz, I. Serano, G. Bueno, and T.-K. Kim, ‘‘Quick violence identification in video,’’ in Proc. Int. Conf. Computer. Vis. Theory Appl. (VISAPP), vol. 2, Jan. 2014, pp. 478–485.

[6] Barett, D.P., Sishkind, J.M.: Activities identification by time sequence of retino topic visual and action features. IEEE Trans. Circuits System. Video Technology. 26(12), 2250–2263 (2015).

[7] Rodriguez , M., et al.: One-shot training of human action with an Map adopted GMM and simplex-HMM. IEEE Trans. Cybern. 47(7), 1769–1780 (2017).

[8] Zhuang, T., et al.: Discriminate dictionary training with action weber localized descriptor for violence identification. IEEE Trans, Circuits System, Video Technology, 27(3), 696–709 (2017).

[9] Wangi, S., et al.: Abnormal activity detection in crowded areas by SL-HOF descriptor and foreground classification. In: 2016 23rd Global Meeting on Pattern Identification. IEEE (2016).

[10] L. Tian , H. Wang, Y. Zhow, and C. Peng , ‘‘Video big-data in smart town: Background construction and utilization for surveillance footage processing,’’ Future Generation. Computer. System., vol. 86, pp. 1371–1382, Sep. 2018.

[11] C. Diman and D. K. Viswakarma, ‘‘A survey of state of the art methods for anomaly human action identification,’’ English. Application. Artifacts. Intelligence., vol. 77, pp. 21–45, Jan. 2018.

[12] P. Zhou , Q. Ding, H. Luoh, and X. Houh, ‘‘Abnormal interaction identification in video based on deep learning techniques,’’ J. Physics., Conference. Series., vol. 844, no. 1, 2017, Art. no. 12044.

[13] S. Chowdhry, M. A. Kaan, and C. Batnagar, ‘‘Several anomalous action identification in videos,’’ Procedia CS., vol. 125, pp. 336–345,Jan. 2018.

[14] H. Lieu, S. Cheng, and N. Kubota , “Intelligent video footages and analytics: A survey,” IEEE Trans. on Industrial Information’s, vol. 9, no. 3, pp. 1222–1233, 2013.

[15] E. B. Nievas, O. Deniz-Su ´ arez, G. B. Garc ´ ´ıa, and R. Sukthankar, “Violence identification in video using computer vision methods,” in CAIP.

[16] L. Xue, C. Gong , J. Yang , Q. Wue, and L. Yaow, “Violent video identification based on MoSIFT attribute and sparse code,” in ICASSP, 2014, pp. 3538–3542.

[17] Y. I. T. Hassner and O. Kliper-Gross, “Violent actions: Real-time recognition of violent crowded activities,” in CVPR workshops, 2012, pp. 1–6.

[18] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu, “Violence recognition using violent actions,” Picture and Video Computing, vol. 48-49, pp. 37–41, 2016.

[19] H. Wang , A. Klasier, C. Schimid, and C. Lieu, “Dense ¨ trajectories and action limited descriptors for action identification,” Global Article of Computer Vision, vol. 103, no. 1, pp. 60–79, 2013.

[20] P. Billinski and F. Bremand, “Human violence identification and classification in surveillance footages,” in AVSS, 2016, pp. 30–36.