ISSN 2347 - 3983
Volume 13. No.4, April 2025
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter011342025.pdf
https://doi.org/10.30534/ijeter/2025/011342025

Identifying Abnormal Human Activity Behavior In Diverse Crowds Using Transfer Learning

Banushri S[1], Dr. Jagadeesha R[2]
[1] Research Scholar, Dept of CSE, ICEAS, VTU-Belagavi, India, banushri914@gmail.com
[2]Professor, Dept of CSE, ICEAS, VTU-Belagavi, India, jagadish.mtech@gmail.com

Received Date: February 26, 2025 Accepted Date: March 27, 2025 Published Date : April 07, 2025

ABSTRACT

Deep learning approaches transform traditional human behavior recognition in video surveillance into more intelligent techniques. This shift offers numerous advanced features that can enhance safety measures in public gatherings. By enabling proactive surveillance, identification, and oversight of diverse crowds, this new paradigm can significantly improve various crowd management operations in terms of effectiveness, capacity, predictability, and safety. Convolutional neural networks (CNNs) present a promising method for recognizing human actions in large datasets, such as those generated by CCTV footage, despite challenges like occlusion, clutter, uneven item distribution, and non-uniform object scale. Therefore, this study proposes a CNN-based Abnormal Classifier model to identify abnormal behavior in public venues such as airports, malls, and theaters. The proposed model classifies behaviors but also delivers definitive results in behavior classification. Experimental results indicate that this method slightly outperforms existing approaches.

Key words: Convolutional neural networks, Classifier, Deep learning, Video surveillance

1. INTRODUCTION

Human abnormal behavior recognition, which depends on video image processing techniques such as image interpretation and visual target tracking, is a recent significant and interesting research topic [1, 2]. It is critical to identify unusual crowd behavior in video surveillance [3-5]. The technique of identifying and classifying certain data, such as density of population and group behavior characteristics, that illustrates unusual crowd behavior from a surveillance image sequence or video is known as abnormal crowd behavior detection [6–8]. The staff must spend observing the surveillance photos to extract valuable data from extensive video recordings and identify unusual behavior and occurrences in the footage [9–11]. Contrarily, manual detection techniques may easily lead to missing and misleading alerts [12]. The term "abnormal behavior detection" refers to a particular frame in a video that can recognize an anomaly and locate the unusual conduct in real-time. Effectively differentiating between normal and abnormal events in the video series requires extracting and classifying pertinent information from the sequence.

In conventional feature extraction approaches, researchers often use temporal and spatial information to predict the kinetic behaviors of visual images. The deep neural network is accurate and successful in voice recognition, image, text, and video processing due to its extensive use and development in business and academia. Deep neural networks are therefore being utilized to address the issue of identifying unusual activity in video. When it involves obscuring moving targets, and targets moving in the same color, the existing algorithm's robustness is compromised.

Security has the utmost emphasis due to the rise in anti-social activities occurring in the area. As a result, organizations need to keep an eye on how individuals connect. Given that it requires ongoing attention, it is nearly impossible for people to monitor. The necessity for an automatic and intelligent interpretation of such video sequences thus presents a difficulty. We propose to tackle this problem with the ability to recognize unusual or anomalous activity on its own.

The current research relates to the area of detecting anomalous crowd behavior. Section 2 discusses related studies. Section 3 explains the suggested methodology, Section 4 provides the results, and the conclusion and recommendations for further work in Section 5.

2. RELATED WORK

The research addressed identifying abnormal behavior detection in videos. The detection of abnormal behavior is reasonably applicable in many areas like traffic violence detection [13] and crowded places like malls, theatres, airports, Railway stations, etc. This section addresses the existing related study done by other researchers.

Authors [14] proposed a model to identify strange conduct using optical flow and Gray Level Co-occurrence Matrix-based text features considering the angle difference between the current frame and the previous frame, abnormal crowd behavior like fright or flee circumstance in which everyone scatters is being detected [15]. Normal event is detected by one class SVM and UMN and PETS2009 datasets were used for the experiment.

Using optical flow volume data and angle change between the following frames, the authors of [16] detected aberrant activity. Then, inputs are fed into CNN for training and classification. From the experiments, the authors found that the Motion information image formation needs more computation time than testing. A research work by [17] used the LK optical flow method for motion estimation and to get precise location information of each instance, texture-based method and entropy were used. However, different illumination conditions are not considered.

To address the problem of the relationship among pictures locally and the limitation of extraction of multiscale features in abnormal behavior detection in a crowd, a model is proposed based on the multiscale network [18]. It increased the performance of the model in complex backgrounds and occlusions. Further, the model can be optimized using parallel CNN to improve intelligent detection. A two-stream network is used to find abnormal behavior in dense crowds using heat maps and optical flows [19]. A new synthetic dataset has also been created to detect the fight and fight unusual behavior. The authors of [20] suggested self-supervised learning based on an Anomalous Event Detection network. This network is composed of PCA-net and k-PCA for anomaly detection. To prevent overfitting, an LRN layer is introduced. [21].

The proposed framework is intended to improve the accuracy of the classifier during the process of bad-lighting videos and videos that have may occlusion images. The proposed model aims to identify the abnormal activity using dense optical flow and two-stream neural networks.

3. PROPOSED METHODOLOGY

The principal aim of the research is to identify and report abnormal activity of people in public places. Figure 1 depicts the workflow of the proposed algorithm.

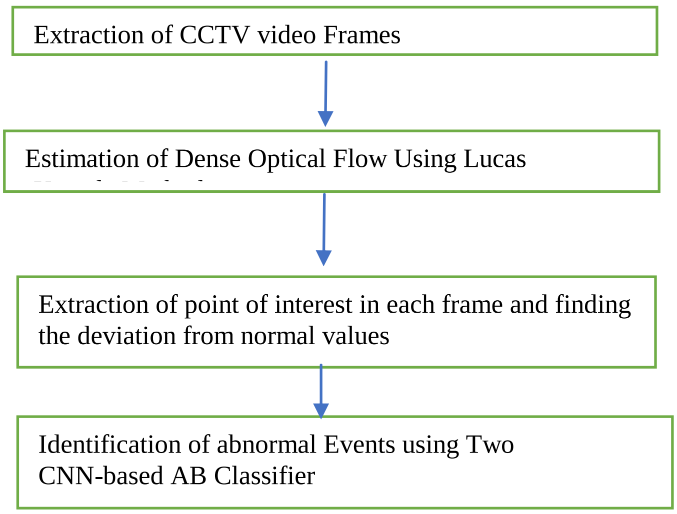

Figure 1: Workflow of the proposed Framework

3.1 Dense Optical Flow Method

To achieve the aforementioned goals, optical flow is a key method for figuring out how picture intensities move and how this motion is related to the movements of objects in the scene [32]. Optical flow is a fairly basic principle used in various ways by most video-processing algorithms. This work analyses the motion vectors' amplitude and orientation to find crowd anomalies. It aids in categorizing the crowd as typical or strange, moving slowly or quickly, and vehicles entering the crowd.

Optical flow estimates the per-pixel motion between two successive frames in a video. In essence, the Optical Flow task suggests that the shift vector for the pixel is to be calculated as an object displacement difference between two adjacent images. Calculating an object's displacement vector motion or camera movements is the fundamental goal of optical flow. The optical flow vector for every pixel in every frame is determined. It starts by locating and extracting the point of interest in each frame. Subsequently, extracted points of interest are tracked frame by frame to determine merging or deviating from the average.

According to the Lucas-Kanade method [33], there will be little to no loss of visual material between two consecutive frames within a certain area of the point p being examined. It follows that every pixel inside a window focused at p is satisfied.

$$\begin{aligned} Imx(pt1)Vx + Imy(pt1)Vy &= -Imt(pt1) \\ Imx(pt2)Vx + Imy(pt2)Vy &= Imt(pt2) \qquad (1) \\ \\ Imx(ptn)Vx + Imy(ptn)Vy &= Imt(ptn) \end{aligned}$$

Where pt1, pt2, pt3….ptn are the pixel points inside the frame window. Im_x (pt_i), $Im_y(pt_i)$, Im_t (pt_i) are the image Im's partial derivatives evaluated at point pt_i and at the current time, about location x, y, and time t. The Lucas-Kanade technique uses the least square principle to arrive at a reasonable response to resolve the 2×2 system.

$$X^TXs = X^T b \text{ or } s = (X^TX)^{-1} X^T b \qquad (2)$$

Matrix X is transposed X^T. At point p, the image's structure tensor is X^TX. The optical flow is computed locally and does not depend on the entire image.

3.2. Architecture of AB Classifier

The proposed architecture applies a two-stream convolutional neural network to enhance the accuracy of unusual activity prediction. Most of the research work used a two-stream CNN for better accuracy. The input data, Spatial-CNN, and motion-CNN [21] extract the spatial and temporal information respectively. A fully connected network is used for classification. This framework outperforms the existing methods. The author of [22] proposed an improved two-stream CNN-based framework to address the problem of overfitting which occurs when the data is imbalanced.

The Dense-Net is used to extract spatial and temporal information and focus loss is applied to address the imbalanced data. [23] proposed a model based on the hybrid two-stream network using CNN and LSTM. Key areas of the human motion footage are extracted using a mixed Gaussian model. The Farneback Dense Option Flow Algorithm is used to obtain spatiotemporal information. The accuracy of the two-stream network with a drop-out mechanism outperforms the existing methods.

The authors of [24] proposed a two-stream inflated 3D CNN (i3d) for unusual behavior detection. It outperformed the current state of techniques in terms of robustness and effectiveness. The two networks in the two-stream i3d extract RGB features from the video and optical flow features respectively. The researchers noted that conventional 3D CNN tends to overlook some motion aspects. So, a two-stream Inflated 3D CNN model was adopted for abnormal behavior detection.

The authors of [24] and [26] used two-stream CNN to improve the accuracy of action recognition in video. Res-Net and Inception-v2 models are applied to extract discrete features and to improve the performance respectively. Many researchers used two-stream CNN models to recognize human behavior in videos. Recent research revealed that the network with deep layers can learn more features in lengthy video and the Residual Network [27] can address the degradation. Hence, as per the recent studies results, our model adopted a two-stream network architecture known as AB classifier for classification by combining space and temporal information.

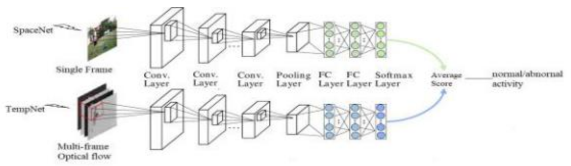

Figure. 2: Architecture of CNN based AB Classifier

Both fully connected and sequential convolutional layers are present in the suggested CNN-based AB classifier. The suggested model architecture is depicted in Figure. 2. Pooling layers are added next to the convolutional layer to reduce the number of features. Two input streams are included to train spatial and temporal information separately and finally fused as the single score for classification.

The definition of layers is as follows: Space- Net generates the class score from a single frame, and Temp- Net generates the class score from multi-frame optical flow. The AB classifier uses the average score for detecting abnormal behavior in video. The AB classifier has two fully connected layers, one pooling layer, and three sequential convolutional layers, and the SoftMax function generates the final score. First, a Convolution Layer of ninety-six 11 X 11 filters of stride of 2. Second, the layer of 256 filters of size 4 X 4, and the number of strides is 4. The third Layer of 512 filters of size 3 X 3 and the number of strides is 2. Then a max pooling layer of 3 X 3 with the number of Strides 2 is added to reduce the inputs and speed up the computation.

The output of the pooling layer is entirely connected to a 4096-node hidden layer, which is fully connected to an additional 4090 hidden layer nodes. The final layer implementing SoftMax contains 500 nodes. This proposed two-stream CNN architecture is implemented with optical flow algorithms for abnormal behavior detection in videos.

3. RESULTS

Experiments were conducted on the Avenue data set [28] and the 3-minute recorded video which contains the normal activity of students sitting in the open ground and the abnormal activity of running in different directions by stamping each other. The Avenue dataset includes the normal activity of pedestrians walking and abnormal activity of people racing, tossing objects, and loitering.

The frames in the data set are resized as 160 x 120 and the video was processed as per the proposed flow. The starting learning rate is 0.001. Additionally, the learning rate drops to 1/10 after every 1000 rounds, and the process ends after 10,000 iterations. TensorFlow and Python software environments are used. The proposed model is compared with the existing method [29] and [30] according to AUC. The experimental results are given in Table 1 and depicted in Figure 3.

Table 1: Comparison of Existing Method

Method type	Accuracy (%) in detecting abnormal activity	No tests conducted
Proposed	93.5	20
The method based on variational autoencoder [30]	93.1	20
The method based on variational autoencoder [29]	82.3	20

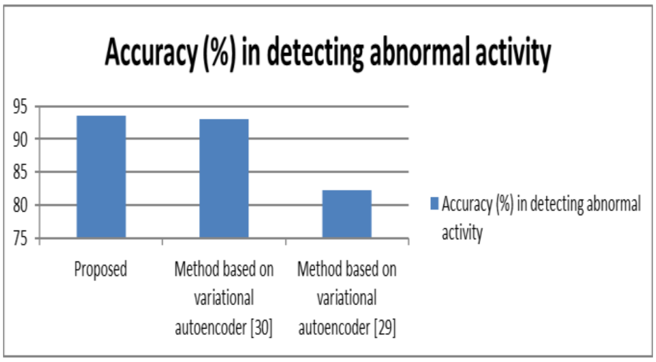

Figure. 3: Comparison of proposed methods with existing method

Regarding frame level AUC, it is evident from Figure. 3 that the suggested strategy performs better. For pixel-level criteria, there is not much difference from the other methods.

5. CONCLUSION

According to recent research studies, detecting abnormal activity in public places is mandatory to avoid unprecedented activities and safeguard humans. Hence, the paper aims to propose the CNN-based model to detect abnormal activity in videos. The experimental output suggested that the proposed model produced superior outcomes. The paper's next steps involve doing a more thorough examination of the suggested model using an intricate dataset.

REFERENCES

1. A. Al-Dhamari, R. Sudirman, and N. H. Mahmood, "**Transfer deep learning along with binary support vector machine for abnormal behavior detection**", IEEE Access, vol. 8, pp. 61085–61095, 2020.

2. X. Hu, J. Dai, Y. Huang et al., "**A weakly supervised framework for abnormal behavior detection and localization in crowded scenes**", Neurocomputing, vol. 383, pp. 270–281, 2020.

3. X. Gao, "**Abnormal behavior detection and warning based on deep intelligent video analysis for geriatric patients**", Journal of Medical Imaging and Health Informatics, vol. 11, no. 2, pp. 345–352, 2021.

4. Q. Cheng, Y. Zhou, Y. Feng, and Z. Liu, "**An unsupervised ensemble framework for node anomaly behavior detection in social network",** Soft Computing, vol. 24, no. 9, pp. 6421–6431, 2020.

5. S. K. Jarraya, M. Masmoudi, and M. Hammami, "**A comparative study of Autistic Children Emotion recognition based on Spatio-Temporal and Deep analysis of facial expressions features during a Meltdown Crisis**", Multimedia Tools and Applications, vol. 80, no. 1, pp. 83–125, 2021.

6. S. Benmakrelouf, C. St-Onge, N. Kara, H. Tout, C. Edstrom, and Y. Lemieux, "**Abnormal behavior detection using resource level to service level metrics mapping in virtualized systems**", Future Generation Computer Systems, vol. 102, pp. 680–700, 2020

7. J. Wang, N. Wang, L. Li, and Z. Ren, "**Real-time behavior detection and judgment of egg breeders based on . Architecture YOLOv3 of Neural CNN based AB Computing and Applications**", vol. 32, no. 10, pp. 5471–5481, 2020.

8. C. Direkoglu, "**Abnormal crowd behavior detection using motion information images and convolutional neural networks,**" IEEE Access, vol. 8, pp. 80408–80416, 2020.

9. J. Hu, X. Zhang, and S. Maybank, "**Abnormal driving detection with normalized driving behavior data: a deep learning approach**," IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 6943–6951, 2020.

10. C.-T. Yang, J.-C. Liu, E. Kristiani, M.-L. Liu, I. You, and G. Pau, "**NetFlow monitoring and cyberattack detection using deep learning with ceph**", IEEE Access, vol. 8, pp. 7842–7850, 2020.

11. O. Ye, J. Deng, Z. Yu, T. Liu, and L. Dong, "**Abnormal event detection via feature expectation subgraph calibrating classification in video surveillance scenes**", IEEE Access, vol. 8, pp. 97564–97575, 2020.

12. T. Wang, M. Qiao, A. Zhu, G. Shan, and H. Snoussi, "**Abnormal event detection via the analysis of multi-frame optical flow information**", Frontiers of Computer Science, vol. 14, no. 2, pp. 304– 313, 2020.

13. S.-Y. Ni, Y.-T. Lan, T. C. I. Lin, and Y.-W. Chen, "**Abnormal behavior detection system and method using the automatic classification of multiple features**", US8885929B2, Nov. 11, 2014, Accessed: Apr. 09,2022.[Online]. Available:

14. h ttps://patents.google.com/patent/US8885929/en

15. R. Lalit and R. K. Purwar, "**Crowd Abnormality Detection Using Optical Flow and GLCM-Based Texture Features**," JITR, vol. 15, no. 1, pp. 1–15, Jan. 2022, doi: 10.4018/JITR.2022010110.

16. C. Direkoglu, M. Sah, and N. E. O'Connor, "**Abnormal crowd behavior detection using novel optical flow-based features,"** in 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Aug. 2017, pp. 1–6. doi: 10.1109/AVSS.2017.8078503.

17. Z. Liu and H. Xia, "**Video Abnormal Behavior Detection Based on Optical Flow Method and Convolutional Neural Network,"** in 2021 International Conference on Human-Machine Interaction, New York, NY, USA: Association for Computing Machinery, 2021, pp. 13–17. Accessed: Apr. 09, 2022. [Online]. Available: https://doi.org/10.1145/3478472.3478476

18. P. Ingole and V. Vyas, "**Anomaly Detection in Crowd Using Optical Flow and Textural Feature,**" in Soft Computing and Signal Processing, Singapore, 2019, pp. 723–732. doi: 10.1007/978-981-13-3600-3_69.

19. B. Song and R. Sheng, **"Crowd Counting and Abnormal Behavior Detection via Multiscale GAN Network Combined with Deep Optical Flow,"** Mathematical Problems in Engineering, vol. 2020, p. e6692257, Dec. 2020, doi: 10.1155/2020/6692257.

20. L. Lazaridis, A. Dimou, and P. Daras, "**Abnormal Behavior Detection in Crowded Scenes Using Density Heatmaps and Optical Flow,**" in 2018 26th European Signal Processing Conference (EUSIPCO), Rome,Sep. 2018, pp. 2060–2064.doi: 10.23919/EUSIPCO.2018.8553620.

21. T. Wang, Z. Miao, Y. Chen, Y. Zhou, G. Shan, and H. Snoussi, "**AED-Net: An Abnormal Event Detection Network,"** Engineering, vol. 5, no. 5, pp. 930–939, Oct. 2019, doi: 10.1016/j.eng.2019.02.008.

22. M. Zhang, C. Gao, Q. Li, L. Wang, and J. Zhang, "**Action detection based on tracklets with the two-stream CNN,"** Multimed Tools Appl, vol. 77, no. 3, pp. 3303–3316, Feb. 2018, doi: 10.1007/s11042-017-5116-9.

23. L. Sha, Y. Zhiwen, X. Kan, Z. Jinli, and D. Honggang, "**An improved two-stream CNN method for abnormal behavior detection,"** J. Phys.: Conf. Ser., vol. 1617, no. 1, p. 012064, Aug. 2020, doi: 10.1088/1742-6596/1617/1/012064.

24. W. Zhang and C. Liu, "**Research on Human Abnormal Behavior Detection Based on Deep Learning**," in 2020 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Jul. 2020, pp. 973–978. doi: 10.1109/ICVRIS51417.2020.00237.

25. P. Jiahui et al., "**An Improved Two-stream Inflated 3D ConvNet for Abnormal Behavior Detection**," Intelligent Automation & Soft Computing, vol. 30, no. 2, p. 673, Aug. 2021, doi: 10.32604/iasc.2021.020240.

26. K. Simonyan and A. Zisserman, **"Two-Stream Convolutional Networks for Action Recognition in Videos**," p. 11, 2014.

27. S. Sarabu and Santra, "**Distinct Two-Stream Convolutional Networks for Human Action Recognition in Videos Using Segment-Based Temporal Modeling**," MDPI, Data, vol. 5, no. 104, pp. 1–12, 2020, doi: http://dx.doi.org/10.3390/data5040104.

28. K. He, X. Zhang, S. Ren, and J. Sun, "**Identity Mappings in Deep Residual Networks**." arXiv, Jul. 25, 2016. doi: 10.48550/arXiv.1603.05027.

29. T. Wang and H. Snoussi, "**Detection of abnormal visual events via global optical flow orientation histogram**," IEEE Transactions on Information Forensics and Security, vol. 9, no. 6, pp. 988–998, 2014.
30. T. Zhao, F. Li, and P. Tian, "**A deep-learning method for device activity detection in mMTC under imperfect CSI based on variational-autoencoder,"** IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 7981–7986, 2020.
31. Qinmin Ma, "**Abnormal Event Detection in Videos Based on Deep Neural Networks**", Scientific Programming, vol. 2021, Article ID 6412608, 8 pages, 2021. https://doi.org/10.1155/2021/6412608.
32. Yuxing Yang, Zeyu Fu, Syed Mohsen Naqvi," **Abnormal event detection for video surveillance using an enhanced two-stream fusion method",** Neurocomputing, Volume 553, 2023, 126561,ISSN 0925-2312.
33. A. K. Alhaq, T. Ahmad and H. Studiawan, "**Forensic Analysis of Car Accident Using MobileNet and Optical Flow,"** 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE), Banda Aceh, Indonesia, 2023, pp. 78-83, doi: 10.1109/COSITE60233.2023.10249437