JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

COMPUTER VISION BASED CCTV ACCIDENT DETECTION USING DEEP LEARNING

DR. P. V KUMAR-Professor, Department of IT, Anurag University,

A. SPANDANA – Student, Master of Computer Applications, Anurag University,

J. SRAVANTHI - Student, Master of Computer Applications, Anurag University,

M. SAI NAGENDRA - Student, Master of Computer Applications, Anurag University,

Abstract

Worldwide information demonstrates that a noteworthy extent of rough fatalities stem from activity collisions. Reaction time for therapeutic help at mishap scenes intensely impacts the probability of survival and is to a great extent affected by human components. Given the predominance of video reconnaissance and cleverly activity frameworks, there is a developing require for robotized strategies to identify activity mischances, especially among computer vision analysts. By and by, Profound Learning (DL) strategies have illustrated impressive viability in complex visual errands, making them alluring for such applications. This think about points to create a DL-based computerized framework for identifying activity mishaps in video film. The proposed approach accept that activity mishap events can be characterized by visual highlights unfurling over time. Thus, the demonstrate engineering comprises of stages for extricating visual highlights and distinguishing transient designs. Visual and transient highlights are learned amid preparing utilizing convolutional and repetitive layers, leveraging both custom and freely accessible datasets. The strategy accomplishes a discovery precision of 98% on open activity mishap datasets, demonstrating strong execution independent of street setups.

1. INTRODUCTION

Different variables contribute to activity mishaps. Among the most predominant causes are street geometry [1], nearby climate conditions [2], inebriated drivers, and over the top speed [3,4]. These mishaps can incur hurt on people included, affecting both activity stream and individual security. Luckily, progressions in innovation have made video observation a profitable instrument for observing and directing urban activity [5]. Be that as it may, the expanding number of required cameras postures challenges in manual observing, requiring robotization instruments. Various methodologies have been proposed to mechanize activity control and checking errands, such as video-based frameworks that assess question speeds and directions [6], pointing to anticipate and anticipate accidents.

The logical community has investigated different approaches for mishap discovery [7], counting statistics-based strategies [8–10], examination of social arrange information [11,12], sensor information utilization [13,14], and the application of machine learning and profound learning [15–18]. Profound learning methods, especially those leveraging convolutional layers in neural systems, have illustrated momentous execution changes in advanced picture preparing errands [19]. These layers abuse spatial connections characteristic in input information, overcoming challenges related with high-dimensional datasets like pictures [22]. Additionally, the integration of different convolutional layers encourages the extraction of pertinent visual highlights, improving organize execution [23–25].

However, certain issues require considering transient connections inside information. Repetitive neural systems (RNNs) address this require by joining associations to past states, empowering them to capture transient conditions [26]. One noteworthy progression in RNNs is the presentation of long short-term memory (LSTM) cells, which exceed expectations in holding both brief- and long-term data [29–31]. In any case, tending to challenges in video investigation requires models that coordinated both convolutional and repetitive layers [32,33]. These half breed models have accomplished advance in assignments like feeling acknowledgment [33], posture estimation [34], sports video examination [35], and activity acknowledgment [36].

Given the complexity of activity mischance location, there is a require to upgrade existing strategies with a novel approach utilizing profound learning procedures for video investigation. Identifying activity mishaps through video examination is computationally seriously and postures a few challenges due to the differing spatio-temporal designs of mishap events. Subsequently, the proposed strategy points to progress execution by successfully leveraging profound learning techniques.

1.2 Scope of the Project:

The venture looks for to progress existing strategies for distinguishing activity mischances by means of video investigation by utilizing progressed profound learning strategies. This includes combining convolutional layers to analyze spatial connections and repetitive layers to comprehend worldly connections. The extend scope envelops looking into existing writing, gathering and planning information, building a profound learning demonstrate, conducting preparing and execution appraisals, joining the show into an operational application, and archiving the whole handle. Through tending to these features, the objective is to set up a strong approach for identifying activity mishaps in real-time, in this manner supporting in the upgrade of street security and activity administration frameworks.

2. LITERATURE SURVEY

The literature survey encompasses a range of studies that delve into the complexities surrounding traffic accidents, exploring various factors contributing to their occurrence and proposing methodologies for their detection and analysis Li (2004) presents a study focused on understanding road traffic dynamics through an urban traffic model of the circular working field. This research likely offers insights into traffic flow patterns and potential bottlenecks, which can be crucial for identifying accident-prone areas and optimizing traffic management strategies. Chu et al. (2019) investigates the intricate relationship between traffic climate, driver behavior, and accident involvement in China. By analyzing these factors, the study aims to uncover underlying patterns and influences that contribute to traffic accidents, providing valuable insights for policymakers and traffic safety initiatives.

Guimarães and da Silva (2019) assess the effectiveness of regulations aimed at controlling alcohol consumption by drivers in the Federal District of Brazil. The study likely evaluates the impact of such regulations on reducing fatal traffic accidents, highlighting the importance of legal measures in promoting road safety. Nishitani (2019) explores the correlation between alcohol consumption and traffic accidents in Japan. By examining this relationship, the study aims to shed light on the role of alcohol impairment in contributing to road traffic incidents, potentially informing targeted interventions and awareness campaigns. Mahata et al. (2019) conducts a spate-temporal analysis of road traffic accidents in large Indian cities. Through this analysis, the study likely identifies spatial and temporal patterns of accidents, offering valuable insights for urban planning and traffic management strategies.

Sheng et al. (2010) proposes a semantic event detection algorithm for traffic surveillance video based on a spate-velocity model. This research likely focuses on leveraging video analytics techniques to detect and classify traffic events, including accidents, thereby enhancing the efficiency of traffic monitoring and management systems. Parsa et al. (2019) apply deep learning techniques to real-time traffic accident detection using spatiotemporal sequential data. By leveraging advanced machine learning algorithms, the study aims to develop more accurate and efficient methods for detecting and predicting traffic accidents, potentially leading to improved road safety measures. Additionally, studies by Joshua and Garber (1990) and Arvin et al. (2019) utilize regression models and connected vehicle message data to estimate accident rates and understand driving behavior at intersections, respectively. These studies offer valuable insights into the statistical modeling of traffic accidents and the utilization of emerging technologies for enhancing traffic safety measures.

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

The current framework basically centers on unmistakable framework or maybe than leveraging Cleverly Transportation Frameworks (ITS) for activity administration, counting blockage and mischance location. Different strategies proposed by analysts include the utilize of smartphones, VANET (Ad-hoc systems), GPS, GSM innovations, and portable applications to naturally identify mischances. Be that as it may, existing frameworks have downsides, such as untrustworthy equipment, especially sensors, and delays in sending mishap alarms due to components like GSM module responsiveness.

3.1.1 Disadvantages of Existing System

Restricted Integration of Brilliantly Transportation Frameworks (ITS): The current framework fundamentally depends on substantial foundation or maybe than completely leveraging Brilliantly Transportation Frameworks (ITS). This confinement prevents the system's capacity to take advantage of progressed innovations for effective activity administration, counting clog and mischance detection.

Reliance on Outside Gadgets: Numerous existing strategies proposed for mishap discovery include the utilize of outside gadgets such as smartphones, VANET (Ad-hoc systems), GPS, GSM advances, and portable applications. This dependence on outside gadgets presents complexities and potential focuses of disappointment, expanding the system's generally defenselessness and diminishing its reliability.

Unreliable Equipment, Especially Sensors: One of the noteworthy disadvantages of the current framework is the lack of quality of equipment components, especially sensors. Breakdowns or mistakes in sensor readings can lead to untrue cautions or missed mischance location, compromising the adequacy of the framework in guaranteeing street safety.

Delays in Mischance Cautions: The framework encounters delays in sending mischance cautions, essentially due to components like the responsiveness of the GSM module. Delays in transmitting mischance data can prevent crisis reaction endeavors and worsen the results of mishaps, expanding the hazard of wounds and fatalities.

Complexity and Support: The complexity presented by the integration of different gadgets and advances in the current framework includes to the upkeep burden. Guaranteeing the legitimate working and synchronization of different components require continuous observing and investigating, devouring assets and time.

Limited Versatility: The dependence on hardware-based arrangements and outside gadgets may constrain the versatility of the current framework. As activity volumes increment or modern regions require checking, scaling up the foundation to suit these changes gets to be challenging and costly.

3.2 Proposed System

The proposed framework presents a novel approach to mischance location utilizing video investigation. Profound learning concepts, counting Convolutional Neural Systems (CNN) with Long Short-Term Memory (LSTM) units, are utilized to prepare a show able of precisely recognizing mishaps in recordings. CNN offers shared-weights engineering and interpretation invariance, upgrading its capacity to analyze visual information successfully.

3.2.1 Advantages of Proposed System

High Accuracy: By utilizing profound learning methods, especially Convolutional Neural Systems (CNN) with Long Short-Term Memory (LSTM) units, the framework can accomplish tall levels of precision in recognizing mishaps in recordings. This empowers more solid recognizable proof of potential occurrences, decreasing wrong alerts and making strides in general framework performance.

Real-time Discovery: The utilize of CNN and LSTM permits for real-time investigation of video film, empowering the framework to identify mischances as they happen. This capability improves reaction times and encourages provoke intercession by specialists or crisis administrations, possibly minimizing the seriousness of mischances and decreasing related risks.

Adaptability: CNNs are known for their capacity to learn and adjust from huge datasets, making them well-suited for analyzing complex visual information like video film. The consolidation of LSTM units assist improves the system's capacity to get it worldly conditions, permitting for vigorous location of mischances over different scenarios and environments.

Efficiency: The shared-weights engineering of CNNs and the memory capabilities of LSTM units contribute to the system's computational effectiveness. This empowers the show to handle expansive volumes of video information productively, making it reasonable for sending in real-world activity administration frameworks without noteworthy computational overhead.

Translation Invariance: CNNs are inalienably translation-invariant, meaning they can distinguish designs and highlights inside pictures notwithstanding of their spatial introduction or position. This property upgrades the system's capacity to distinguish mishaps over diverse camera points and points of view, moving forward its generally viability in different activity checking scenarios.

Reduced Reliance on Equipment: Not at all like conventional frameworks that depend intensely on physical sensors and equipment components, the proposed framework fundamentally leverages software-based profound learning calculations. This diminishes reliance on equipment foundation, possibly bringing down usage costs and rearranging framework upkeep and scalability.

3.3 Proposed System Design

In this project work, there are three modules and each module has specific functions, they are:

1. DATA COLLECTION
2. PRE-PROCESSING
3. ACCIDENT DETECTION

3.3.1 DATA COLLECTION

For the solution proposed, image data sets. The first one consists of images used for the fine tuning of the visual feature vector extractor. The image dataset was built from scratch, applying the web scraping technique to populate the dataset. For this, a series of logical steps were proposed. First, we identified the sources on the web where the image search was performed. Next, we defined the set of keywords for the searches. For this process, the following keywords were selected: Traffic accidents, Car accidents, Motorcycle accidents, and Truck accidents. Then, the automation stage was performed. The application was developed in the Python programming language together with the Selenium library, which contains useful functions to perform this process. Finally, a manual validation of all the collected images was carried out together with an image transformation in order to standardize the size and format used.

3.3.2 PRE-PROCESSING

A image is segmented in order to obtain a greater number of examples with a certain number of constant images and, in turn, a segment with shorter duration. This is because traffic accidents have a short average duration (10 frames) , which allows for processing of the original video in a more efficient way. In order to select the segmentation technique for the input data, some experiments were performed on the videos taken from the dataset. The techniques to be evaluated were compared using the same videos in each case. The first technique consists of a segmentation without frame discrimination. Therefore, all consecutive images of the video are selected until the maximum time of the segment is reached. This technique has an average reading time of 0.18 s. The second technique used seeks to skip frames in order to reduce the redundancy that can be observed when using very close images in the video. This is because when the video has been recorded with a traditional camera, the number of similar consecutive frames is very high. For this reason, we experimented by skipping one frame for each frame selected. That is, in this case, the images with an odd index were chosen from the video, until the maximum length of duration established for the segment was reached.

3.3.3 ACCIDENT DETECTION

The solution presented is based on a visual and a temporal feature extractor. The first stage of the model consists of the PYTORCH architecture (pre-trained with the accident dataset). That is, all the Inception cells (convolutional layers) were used, eliminating the multilayer perceptron at the end of this architecture. This is to use this part of the model only as a visual feature extractor, upper part. However, by performing multiple experiments, it was concluded that the pre-trained model does not differentiate between a vehicle at rest and a vehicle hit by a traffic accident. Therefore, the images dataset was used for training in order to adjust the weights of this pre-trained network. In this process, all the weights of the initial layers of the architecture were frozen, and only those of the last convolutional cell of pytorch were adjusted. To adjust the feature extractor, multiple experiments were performed.

3.4 Architecture

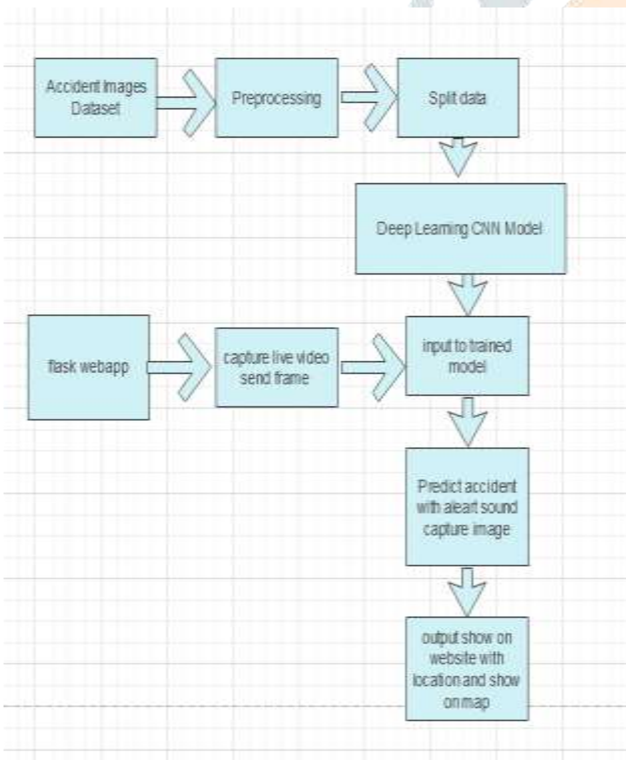
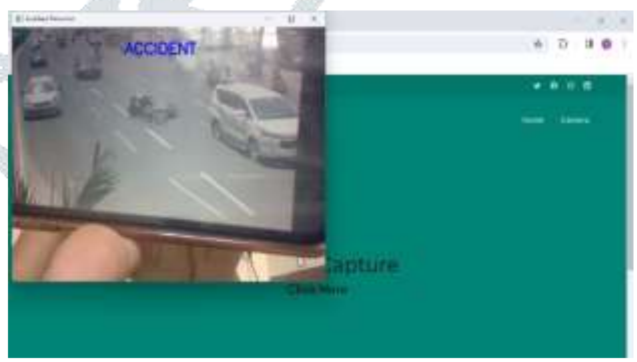
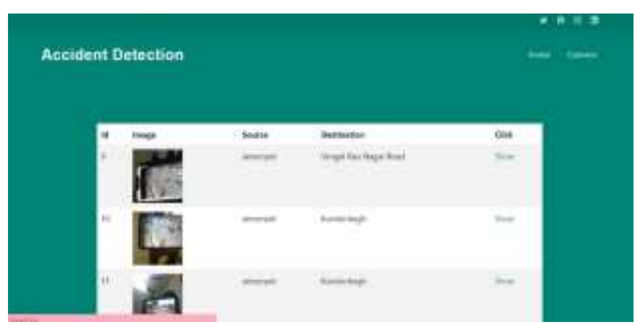


Fig 1: System Architecture

4. RESULT SCREEN SHOTS





5. CONCLUSION

Pre-trained neural systems are not able to compute a vector with important highlights for exceptionally particular issues. Hence, it is fundamental to alter the weights of these models' utilizing cases of the issue to be unraveled. The procedure that best speaks to a worldly fragment of a activity mischance does not dispense with any information, since the closeness values between the portions of the procedures with outline choice show insignificant contrasts between them, whereas the computational fetched, preparing time and precision in mishap discovery show superior comes about by not conditioning the determination of outlines to a metric. Manufactured vision has made extraordinary propels in the understanding of video scenes. One of the best-performing methods is manufactured neural systems. Numerous of these models are based on structures composed of convolutional layers and repetitive layers, in arrange to extricate as much data as conceivable from the input information. The proposed strategy is based on this sort of design and accomplishes a tall execution when identifying activity mishaps in recordings, accomplishing an F1 score of 0.98 and an precision of 98%. The proposed show appears tall execution for video activity mishap discovery. In any case, due to the scarcity of such datasets in the logical community, the conditions beneath which the show works are constrained. The arrangement is confined to vehicular collisions, barring cruisers, bikes, and people on foot due to the insignificant number of these sorts of illustrations accessible. In expansion, the show has mistakes in deciding mishap sections with moo light (such as nighttime recordings) or moo determination and impediment (moo quality video cameras and areas).

6. REFERENCES

[1] Li, M.Z. "The Road Traffic Analysis Based on an Urban Traffic Model of the Circular Working Field." Acta Mathematica Applicata Sinica 2004, 20, 77–84.

[2] Chu, W.; Wu, C.; Atombo, C.; Zhang, H.; Özkan, T. "Traffic Climate, Driver Behaviour, and Accidents Involvement in China." Accident Analysis & Prevention 2019, 122, 119–126.

[3] Guimarães, A.G.; da Silva, A.R. "Impact of Regulations to Control Alcohol Consumption by Drivers: An Assessment of Reduction in Fatal Traffic Accident Numbers in the Federal District, Brazil." Accident Analysis & Prevention 2019, 127, 110–117.

[4] Nishitani, Y. "Alcohol and Traffic Accidents in Japan." IATSS Research 2019, 43, 79–83.

[5] Mahata, D.; Narzary, P.K.; Govil, D. "Spatio-Temporal Analysis of Road Traffic Accidents in Indian Large Cities." Clinical Epidemiology and Global Health 2019, 7, 586–591.

[6] Sheng, H.; Zhao, H.; Huang, J.; Li, N. "A Spatio-Velocity Model Based Semantic Event Detection Algorithm for Traffic Surveillance Video." Science China Technological Sciences 2010, 53, 120–125.

[7] Parsa, A.B.; Chauhan, R.S.; Taghipour, H.; Derrible, S.; Mohammadian, A. "Applying Deep Learning to Detect Traffic Accidents in Real Time Using Spatiotemporal Sequential Data." arXiv 2019, arXiv:1912.06991.

[8] Joshua, S.C.; Garber, N.J. "Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression Models." Transportation Planning and Technology 1990, 15, 41–58.

[9] Arvin, R.; Kamrani, M.; Khattak, A.J. "How Instantaneous Driving Behavior Contributes to Crashes at Intersections: Extracting Useful Information from Connected Vehicle Message Data." Accident Analysis & Prevention 2019, 127, 118–133.