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

ACCIDENT DETECTION USING DEEP LEARNING

Sai Keertana S - Student, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Gopichand Reddy T- Student, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Vardhan Reddy V- Student, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Anusha Sri M - Assistant Professor, Department of CSE – Data Science, Geethanjali College of Engineering and Technology

Abstract

According to worldwide statistics, traffic accidents are the cause of a high percentage of violent deaths. The time taken to send the medical response to the accident site is largely affected by the human factor and correlates with survival probability. Due to this and the wide use of video surveillance and intelligent traffic systems, an automated traffic accident detection approach becomes desirable for computer vision researchers. Nowadays, Deep Learning (DL)-based approaches have shown high performance in computer vision tasks that involve a complex features relationship. Therefore, this work develops an automated DL-based method capable of detecting traffic accidents on video. The proposed method assumes that traffic accident events are described by visual features occurring through a temporal way. Therefore, a visual features extraction phase, followed by a temporary pattern identification, compose the model architecture. The visual and temporal features are learned in the training phase through convolution and recurrent layers using built-from-scratch and public datasets. An accuracy of 98% is achieved in the detection of accidents in public traffic accident datasets, showing a high capacity in detection independent of the road structure.

1. INTRODUCTION

There are different factors that cause traffic accidents. Among the most common factors that increase the probability of their occurrence are the geometry of the road [1], the climate of the area [2], drunk drivers, and speeding [3,4]. These accidents can cause harm to the people involved and, although most of these present only material damage, each one affects people's quality of life in terms of both traffic mobility and personal safety. Thanks to technological advances, video cameras have become a resource for controlling and regulating traffic in urban areas. They make it possible to analyze and monitor the traffic flowing within the city [5]. However, the number of cameras needed to perform these tasks has been increasing significantly over time, which makes control difficult if automation mechanisms are not implemented because the number of professionals needed to comply with all the points also increases. Several approaches have been proposed to automate tasks within the control and follow-up process. An example of this is a system based on video camera surveillance in traffic. Through these, it is possible to estimate the speeds and trajectories of the objects of interest [6], with the objective of predicting and controlling the occurrence of traffic accidents in the area. The scientific community has presented different approaches to detect traffic accidents [7]. These include statistics-based methods [8–10], social network data analysis [11,12], sensor data [13,14], machine learning, and deep learning [15–18]. These latest techniques have presented improvements in various fields of science, including video-based problem solving (video processing). Therefore, it is important to study these tech.

niques in order to approach a solution to the detection and classification of traffic accidents based on video. With the advent of convolutional layers in the domain of neural networks, better performance has been achieved in the solution of problems involving digital image processing [19]. Deep learning techniques have shown high performance in a large number of problems, especially for image understanding and analysis [20,21]. These layers exploit the spatial relationship that the input data possess and that, due to the size of the information, it is not possible to achieve with dense neural networks [22]. The use of convolutions on input data with a large number of features makes it possible, among other things, to avoid the problem of the curse of dimensionality. This is a very frequent problem when working with data with high complexity, such as images. Likewise, it is important to highlight that the use of several convolutional layers helps the extraction of relevant visual features within the same dataset, which defines the performance of the network [23–25]. On the other hand, there are problems where the spatial relationship of the data is not a determining characteristic. In some problems, the temporal relationship that the data may have is of greater importance. This is because there are events that depend on past and/or future events, that is, on a context of the event in time in order to understand the real event. This is why a new deep learning model has emerged: recurrent neural networks. These networks have a similar architecture to dense artificial neural networks but differ in that at least one neuron has a connection to itself. This allows them to be able to remember what has been previously processed, i.e., it gives them the ability to store information over periods of time (data memory).

They specialize in finding the temporal relationships that a set of data may have. Such networks are used to solve problems such as rate-of-change prediction [26], text translation [27], and natural language processing [28], among others. The data processing in these neurons has a higher complexity than the processing performed from a traditional neuron. In addition, these have been improved over the years. One of the most relevant changes was the possibility that the cell can store short- and long-term memory, called long short-term memory neurons (LSTM). These networks have presented improvements in several problems with respect to past models. Among these are travel time prediction problems [29], language understanding [30], and natural language processing [31]. However, the analysis of video scenes is not a problem that can be solved using one of the two models mentioned above. This is because a video presents both a spatial and a temporal relationship in its content. Therefore, the scientific community has presented several architectures that use both deep learning layers: convolutional layers and recurrent layers [32,33]. Some of the advances they have achieved using these types of architectures are emotion recognition [33], estimation of a person's posture [34], analysis of basketball videos for the automation of tasks such as the score of each team [35], and action recognition [36]. Because of this, a method capable of solving the traffic accident detection problem is proposed. However, the process of detecting traffic accidents is a task that involves a lot of processing and, for this reason, these tasks present many difficulties. The occurrence of a road accident is an event capable of occurring in multiple patio-temporal combinations. This leaves a large domain of diverse distributions of data to be classified as an accident, which makes it difficult to solve the problem. Similarly, the classification of an accident is a complex problem due to the temporal implications it may present. Therefore, we seek to improve the performance of current approaches with the design of a method capable of detecting traffic accidents through video analysis using deep learning techniques.

1.2 Scope of the Project:

The project aims to enhance current approaches to detecting traffic accidents through video analysis by leveraging deep learning techniques, particularly combining convolutional layers for spatial relationship analysis and recurrent layers for temporal relationship understanding. The scope includes conducting a literature review, collecting and preprocessing data, developing a deep learning model, training and evaluating its performance, deploying and integrating the model into a practical application, and documenting the entire process. By addressing these aspects, the goal is to create a robust method for real-time traffic accident detection, contributing to improved road safety and traffic management systems.

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 system primarily focuses on tangible infrastructure rather than leveraging Intelligent Transportation Systems (ITS) for traffic management, including congestion and accident detection. Various methods proposed by researchers involve the use of smartphones, VANET (Ad-hoc networks), GPS, GSM technologies, and mobile applications to automatically detect accidents. However, existing systems have drawbacks, such as unreliable hardware, particularly sensors, and delays in sending accident alerts due to factors like GSM module responsiveness.

3.1.1 Disadvantages of Existing System

Limited Integration of Intelligent Transportation Systems (ITS): The current system primarily relies on tangible infrastructure rather than fully leveraging Intelligent Transportation Systems (ITS). This limitation hinders the system's ability to take advantage of advanced technologies for efficient traffic management, including congestion and accident detection.

Reliance on External Devices: Many existing methods proposed for accident detection involve the use of external devices such as smartphones, VANET (Ad-hoc networks), GPS, GSM technologies, and mobile applications. This reliance on external devices introduces complexities and potential points of failure, increasing the system's overall vulnerability and reducing its reliability.

Unreliable Hardware, Particularly Sensors: One of the significant drawbacks of the current system is the unreliability of hardware components, particularly sensors. Malfunctions or inaccuracies in sensor readings can lead to false alarms or missed accident detections, compromising the effectiveness of the system in ensuring road safety.

Delays in Accident Alerts: The system experiences delays in sending accident alerts, primarily due to factors like the responsiveness of the GSM module. Delays in transmitting accident information can hinder emergency response efforts and exacerbate the consequences of accidents, increasing the risk of injuries and fatalities.

Complexity and Maintenance: The complexity introduced by the integration of multiple devices and technologies in the current system adds to the maintenance burden. Ensuring the proper functioning and synchronization of various components require ongoing monitoring and troubleshooting, consuming resources and time.

Limited Scalability: The reliance on hardware-based solutions and external devices may limit the scalability of the current system. As traffic volumes increase or new areas require monitoring, scaling up the infrastructure to accommodate these changes becomes challenging and costly.

3.2 Proposed System

The proposed system introduces a novel approach to accident detection using video analysis. Deep learning concepts, including Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) units, are employed to train a model capable of accurately detecting accidents in videos. CNN offers shared-weights architecture and translation invariance, enhancing its ability to analyze visual data effectively.

3.2.1 Advantages of Proposed System

High Accuracy: By utilizing deep learning techniques, particularly Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) units, the system can achieve high levels of accuracy in detecting accidents in videos. This enables more reliable identification of potential incidents, reducing false alarms and improving overall system performance.

Real-time Detection: The use of CNN and LSTM allows for real-time analysis of video footage, enabling the system to detect accidents as they occur. This capability enhances response times and facilitates prompt intervention by authorities or emergency services, potentially minimizing the severity of accidents and reducing associated risks.

Adaptability: CNNs are known for their ability to learn and adapt from large datasets, making them well-suited for analyzing complex visual data like video footage. The incorporation of LSTM units further enhances the system's ability to understand temporal dependencies, allowing for robust detection of accidents across various scenarios and environments.

Efficiency: The shared-weights architecture of CNNs and the memory capabilities of LSTM units contribute to the system's computational efficiency. This enables the model to process large volumes of video data efficiently, making it suitable for deployment in real-world traffic management systems without significant computational overhead.

Translation Invariance: CNNs are inherently translation-invariant, meaning they can detect patterns and features within images regardless of their spatial orientation or position. This property enhances the system's ability to detect accidents across different camera angles and perspectives, improving its overall effectiveness in diverse traffic monitoring scenarios.

Reduced Dependency on Hardware: Unlike traditional systems that rely heavily on physical sensors and hardware components, the proposed system primarily leverages software-based deep learning algorithms. This reduces dependency on hardware infrastructure, potentially lowering implementation costs and simplifying system maintenance 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

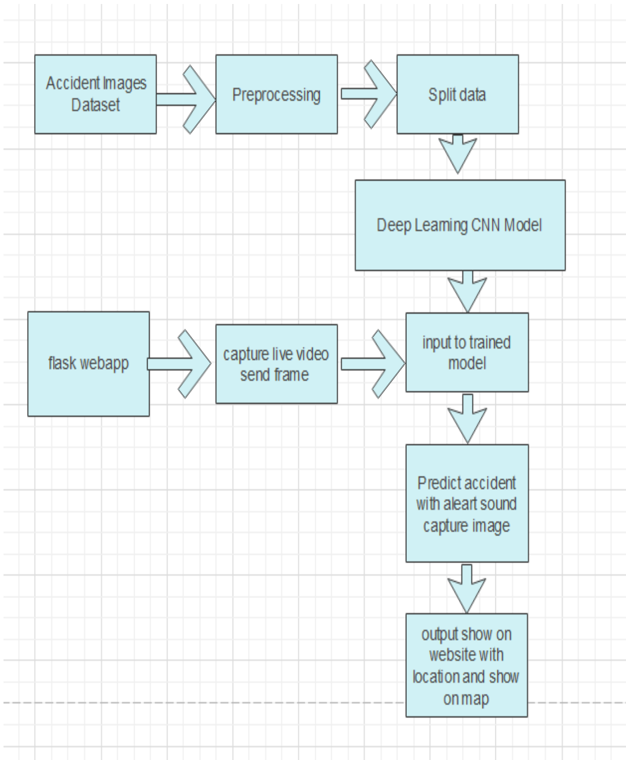
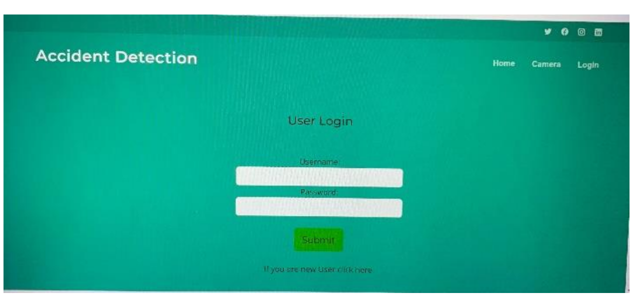
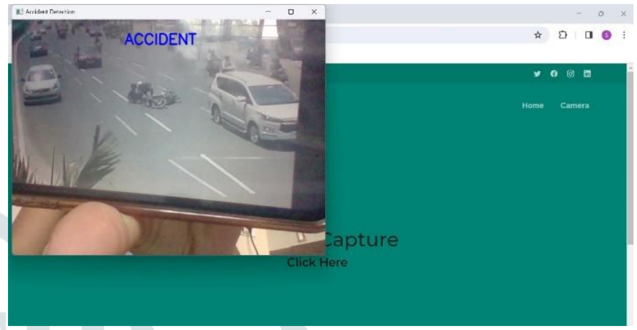
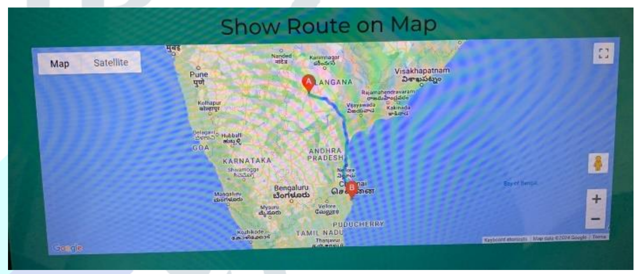


Fig 1: System Architecture

4. RESULT SCREEN SHOTS







5. CONCLUSION

Pre-trained neural networks are not able to compute a vector with relevant features for very specific problems. Therefore, it is necessary to adjust the weights of these models' using examples of the problem to be solved. The technique that best represents a temporal segment of a traffic accident does not eliminate any data, because the similarity values between the segments of the techniques with frame selection present negligible differences between them, while the computational cost, processing time and accuracy in accident detection present better results by not conditioning the selection of frames to a metric. Artificial vision has made great advances in the understanding of video scenes. One of the best-performing techniques is artificial neural networks. Many of these models are based on architectures composed of convolutional layers and recurrent layers, in order to extract as much information as possible from the input data. The proposed method is based on this type of architecture and achieves a high performance when detecting traffic accidents in videos, achieving an F1 score of 0.98 and an accuracy of 98%. The proposed model shows high performance for video traffic accident detection. However, due to the paucity of such datasets in the scientific community, the conditions under which the model works are limited. The solution is restricted to vehicular collisions, excluding motorcycles, bicycles, and pedestrians due to the negligible number of these types of examples available. In addition, the model has errors in determining accident segments with low illumination (such as nighttime videos) or low resolution and occlusion (low quality video cameras and locations).

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.