EARS(Emergency Alert Response System) for Accident Detection using Real time Image processing and Video processing.

Sneha Dhondugade[1], Seema Kale[2], Pradnya Bhor[3], Zaid Khan[4], Siddharth Shrivastav[5], Prof. Rinku A. Badgujar[6].

[1,2,3,4,5]JSPM's BSIOTR, Wagholi, Pune.
[6] Professor, Dept. of Computer Engineering, JSPM's BSIOTR, Wagholi, Pune.
sneha.dhondugade3@gmail.com

Abstract

Nowadays road accident is the most unwanted thing, but they happen quite often in day to day life to a road user. It happens mostly because of human errors i.e. Over speeding, Drunken driving, etc. It can cause serious injuries that may lead to lifelong disabilities and even death, each minute matters when it is difference in between life and death. So, rapid assistance to injured people should be provided as early as possible. Many cities in India are equipped with video surveillance cameras installed on different roads and highways. In this research, we are processing these CCTV footage to detect accident using various algorithms like CNN and RNN, and will notify to respective responsible authorities for further help. System extracts frames from video and does image classification by processing them. CNN classifies single frame at a time whereas RNN detects and classifies multiple frames at a time concurrently. By considering threshold values of factors like variation rate of the velocity, rate of noise generated as well as position, area, direction of moving vehicles, system will recognize that accident has occurred or not. If accident has been occured information regarding area, accident status is sent to respective police and hospital of that particular area. Hence, rapid assistance will be provided to distressed people immediately. By applying Data Analytics on the gathered information comparison will be done and reports will be generated.

Keywords: Accident, CNN, RNN, Image processing, Image classification, Video, Classification, Data Analytics.

1. Introduction

In most of the cases, accidents occur because of collision of vehicles with other vehicles, pedestrians or objects. Accidents commonly tend to occur at Road junctions and experience crashes due to angle and left-right turns. So, accurate detection of accidents at Road Junctions offers tremendous benefits of saving properties and lives and minimizing congestion and delay.

Accident detection includes processing videos which has been taken as a input from CCTV and classifying them as accident happened or not and notifying to respective prestige. There are various methods existing that use fuzzy theory and image-processing techniques for prediction of incident before it happens. By using "behavioral abnormality" of continuous frames the judgment of whether an incident has happened or not is made. Unfortunately, these methods does not work at the road junction as they are complicated places. Hence, we suggested various algorithms for image and video processing as well as classification for automatic detection of accident.

The goal of this current research is to develop an automatic accident detection system, which can process videos recorded from stationary cameras over roads e.g. CCTV cameras installed near traffic junctions / intersections and process them using Convolution Neural Networks(CNN) and Recurrent Neural Networks(RNN) Algorithms. CNN is generally used for image recognition and object classification. It usually takes image as a 2D matrix, process it and classify it under certain categories. It detects and classifies single frame at a time. Whereas RNN is next step of CNN. RNN detects and classifies multiple frames at a time

concurrently. It has LSTM(Long Short Term Memory) for storing previous videos to analyze based on dataset. RNN can handle sequential data while CNN cannot.

2. Related Work

In the recent years of researches, various approaches have been applied in this particular area of detecting accident but still the gap is there as it needs improvement in detection and tracking for accurate prediction. Yong-Kul Ki [2] applied the technique of vision-based traffic accident detection system for automatically detecting, recording, and reporting traffic accidents at intersections. This model first extracts the vehicles from the video image of CCD camera, tracks the moving vehicles, and extracts features such as the variation rate of the velocity, position, area, and direction of moving vehicles. But this is efficient in case of image only, it means it requires exact image of collision, it does not work for video.

However, the performance of this system is significantly affected by the selected thresholds. A vision based detection and attribute-based search of vehicles in dense traffic monitoring has been presented earlier using multiple detectors and can be extended to large scale adaptation. Moving object detection algorithm from video was also proposed earlier for localization of vehicle by differentiating current image and background image and applying connectivity and relabeling technique to detect vehicles. Although the approach has filtered background noise from video using opening operation, still it has some noise clustering which cannot be filtered easily [3].

3. Proposed System

The proposed EARS(Emergency Alert Response System) for accident detection makes use of video data acquired from stationary traffic cameras, performing causal mathematical operations over a set of frames obtained from the video to estimate the prediction of an event. CNN and RNN algorithms are applied on each frame that are extracted from video.

Every real-world image can be annotated with multiple labels, because an image normally contains rich semantic information, such as objects, parts, scenes, actions, and their interactions or attributes. Modeling the rich semantic information and their dependencies is essential for image understanding. So, multi-label classification of image is receiving increasing attention.

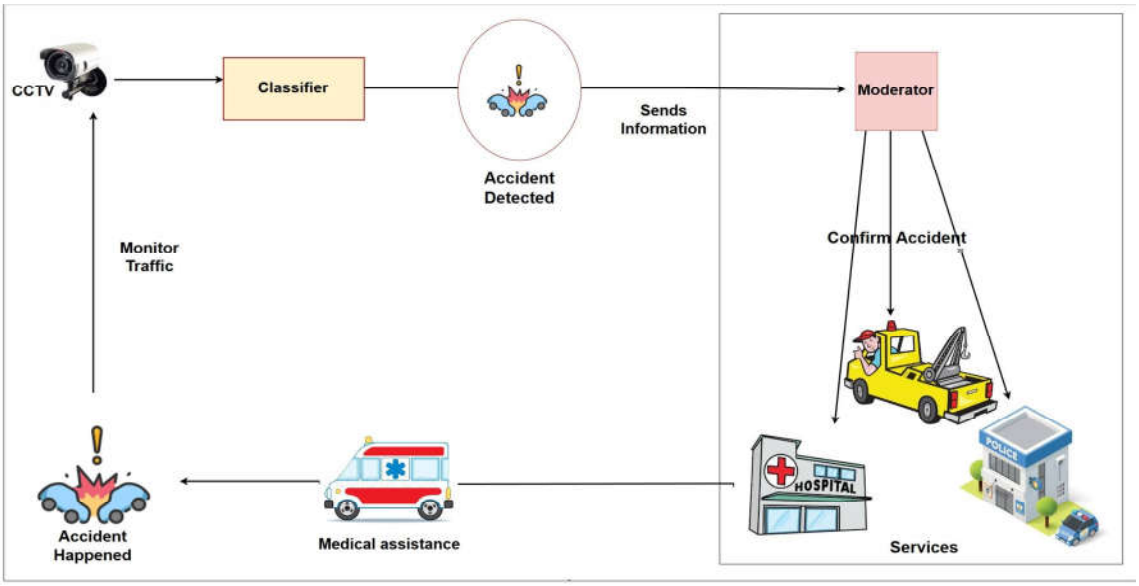

Fig 1 - EARS Architecture.

By comparing with various mathematical models frames will be classified accordingly and accident will be detected. If accident has been occurred then it will be notified to moderator system and information regarding accident will be stored in database. Information contains accident information, site condition, cause of the accident, vehicles, and driving pattern at accident time. Moderator system will try to recognize the nearby hospital and police station by comparing with received information about accident. Thus, rapid assistance will be provided to distressed people.

4. Accident Detection By Combination Of Mlp Classifiers

Accident detection algorithm generally consist of three steps i.e. vehicle extraction, feature extraction of a moving vehicle (MV), and accident detection. A Multilayer Perceptron(MLP) is class of feedforward artificial neural network which consist of at least 3 layer nodes : an input layer, a hidden layer and an output layer. We suggest a CNN-RNN framework for multi label classification problem. The illustration of the CNN and RNN is shown below. It mainly contains two parts - The CNN part extracts semantic representations from images and the RNN part models image/label relationship and label dependency.

4.1) Vehicle Extraction And Tracking

As a video is nothing but collection of continuous frames from which vehicles are extracted by detecting moving parts in each frame based on a difference equation. This process consists of taking the difference of two consecutive frames, binarization, and horizontal and vertical projection, and then extracting parts which exceed the threshold value. For the extraction of moving regions in a video sequence, an input image, a pair of gray-level images, Ik-1 (x, y) and Ik (x, y) acquired at successive time instants τk-1 and τk, respectively[2]. The output is nothing but moving regions in which significant changes have been detected. For the extraction of moving regions, the difference image, D(x, y) is computed which is as follows :

$$D(x, y) = Ik(x, y) – Ik\text{-}1(x, y).$$

4.2) Convolution Neural Network

Convolution Neural Network has long been used in the field of digital image processing and speech recognition, and has achieved great success in various fields. Before the convolutional neural network was proposed, image processing were done by traditional machine learning algorithms. Although great results were achieved, it was difficult to make further breakthroughs, so CNN came into being use. Currently, CNN for image processing is relatively mature.

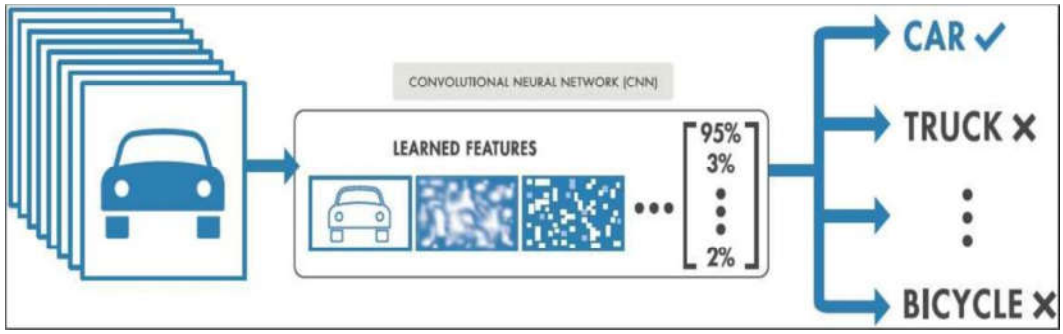

Figure 2 : CNN - Image Recognition.

As shown in the figure 2, image recognition task which is performed by CNN, there is only one object to be recognized in the center of the general image, and the task is to determine what the image is. But in reality, when we look around the world with our eyes, we carry out far more complex tasks. CNN imports a picture as it is, processes the image, using the box to correctly identify where the main object is in the image. Now that we have found the object in the box, we can narrow the border to make it more in line with the object's 3D dimensions[1]. The box is then divided into small boxes as input(training dataset) to train three

different models: a CNN that generates image features, a classifier that predicts categories, and a regression model that tightens borders.

4.3) Recurrent Neural Network

CNN have shown a great success in single-label image classification, it is important to note that real world images generally contain multiple labels, which could correspond to different objects, scenes, actions and attributes in an image. Since we aim to characterize the high-order label correlation, we employ Long Short Term Memory (LSTM) neurons as our recurrent neurons, which has been demonstrated to be a powerful model of long-term dependency.

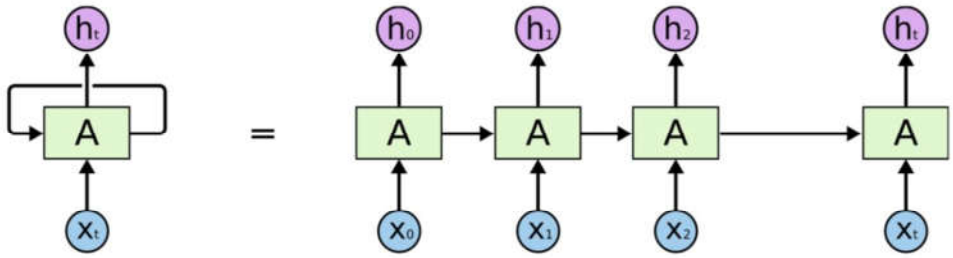

Figure 3 : RNN Model

RNN is a class of neural network that maintains internal hidden states to model the dynamic temporal behaviour of sequences. LSTM extends RNN by adding three gates to an RNN neuron: a forget gate f to control whether to forget the current state; an input gate i to indicate if it should read the input; an output gate o to control whether to output the state[6]. These gates enable LSTM to learn long-term dependency in a sequence, and make it is easier to optimize, because these gates help the input signal to effectively propagate through the recurrent hidden states r(t) without affecting the output[6].

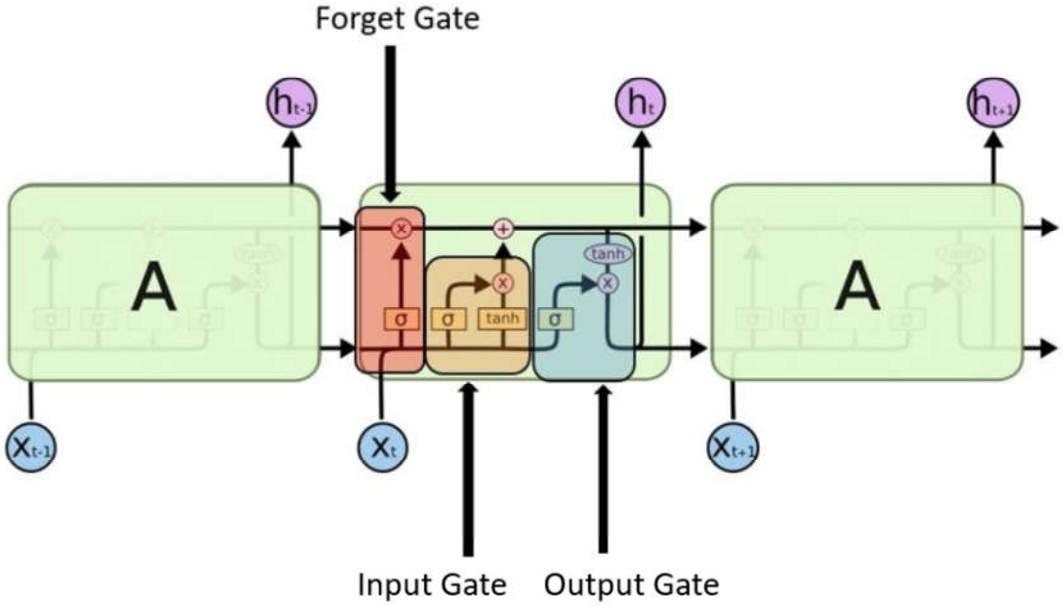

Figure 4 : RNN - LSTM.

5) Conclusion

We have demonstrated a promising approach for an image and video processing system for automatically detection, recording, and reporting accidents at an intersections. And we suggested to to design the metadata registry to maintain information regarding accident and so that we can perform various data analytics operations for future safety as well. EARS will help the people suffered from accident by notifying to respective police stations and hospitals, so that there valuable life can be saved. It will help in reducing number of accidents and traffic as well. It will play a vital role in infrastructure development. Videos captured by cameras can be used as evidence of accident and can be used to avoid accident in future. This data will be useful in future for analytical references.

6) References

[1] Yong-Kul Ki†. "Accident Detection System using Image Processing and MDR". VOL.7 No.3, March 2007.

[2] JIUDONG YANG, JIANPING LI. "APPLICATION OF DEEP CONVOLUTION NEURAL NETWORK". 978-1-5386-1010-7/17/$31.00 2017

[3] Prem Kumar Bhaskar, Suet- Peng Yong. "Image Processing Based Vehicle Detection and Tracking Method". 978-1-4799-0059-6/13/$31.00 2014.

[4] JiangWang, Yi Yang, JunhuaMao, Zhiheng Huang, ChangHuang, Wei Xu. "CNN-RNN : A Unified Framework for Multi-label Image Classification".

[5] Lilia Kechiche, Lamjed Touil, Bouraoui ouni. "Real-time Image and Video Processing: Method and architecture" 978-1-4673-8526-8/16/$31.00 2016.