Prediction of Pedestrian Road Crossing Intention Using Deep Learning: A Review

Rohit A. Telkar [1], Shubham A. Thombare [1], Mayur P. Zagade [1], Omkar K. Kshirsagar [1]

[1]*Student, Dept. of Information Technology, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, India.*

Corresponding Author: rohittelkar0581@gmail.com

Abstract: - Nowadays, Pedestrian are the most vulnerable road users which are highly prone to road accidents with vehicles. Pedestrian road crossing intention recognition plays a crucial role in ensuring the safety of both pedestrians and drivers. In Advanced Driver Assistance System (ADAS) in vehicles, Detection and Prediction is the challenging task in this system we proposed advanced Deep Learning and Computer vision-based model for pedestrian detection and prediction. A You Only Look Once (YO-LO) algorithm is used for detection of pedestrian. Convolutional Neural Network (CNN) used to predict the intention of pedestrian. JAAD Dataset is used for system training and testing.

Key Words: - ***Deep Learning, Convolutional neural Network (CNN), You Only Look Once (YOLO), Advanced Driver Assistance System (ADAS).***

I. INTRODUCTION

In last one year, motor vehicle crashes on roadways resulted in the deaths of more than 7,000 pedestrians. An average of one death occurs every 75 minutes. Pedestrians made up one in every six fatal accident victims. 1–2 A total of 104,000 pedestrians were treated in emergency rooms for non-fatal crash-related injuries throughout the course of the year. A high-tech automobile feature called pedestrian detection works to safeguard both passengers and bystanders.

In order to identify human movement, pedestrian detection systems employ a variety of cameras positioned all around the vehicle, together with computers and sensors.

Manuscript revised May 21, 2023; accepted May 22, 2023. Date of publication May 25, 2023.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

Technology advancements have made it simple to tell people apart from other stationary things you could see on the road. Some vehicles can use this information to automatically apply the brakes when there is a possibility of running over a pedestrian. Low speeds are ideal for the system since they give the automobile enough time to slow down and reduce the likelihood of a collision. Even though roads are a necessary component of daily commutes, they are nonetheless dangerous for commuters, especially for pedestrians. Because of the background's extreme changeability and significant overlapping or occlusion, pedestrian recognition remains one of the most challenging areas of object detection.

Since pedestrian recognition should be precise as well as quick enough to be used in real time, achieving high accuracy and real time speed also poses a significant difficulty. pedestrian intention of crosswalks that are accurate could reduce the number of injuries caused by traffic and increase pedestrian safety. For pedestrian detection, a You Only Look Once (YO-LO) [1] algorithm can be utilised. The intention of a pedestrian is predicted using a convolutional neural network (CNN) in conjunction with a gated recurrent unit (GRU) [1] and an attention mechanism. JAAD Dataset is used for system testing and training.

II. LITERATURE SURVEY

The authors R. Quan et al proposes a long-short-term memory (LSTM) for adaptively integrating multiple sources of information from pedestrians and vehicles. Firstly, introducing additional memory cells to improve the portability of LSTMs when modelling future variations. These additional storage cells include a speed cell for explicitly modelling vehicle speed dynamics, an intent cell for dynamically analysing pedestrian crossing intent, and a of correlated cells are included. These three separate cells reveal the future motion of vehicles, pedestrians, and the global scene. Second, proposed a gated shift maneuverer that learns pedestrian movements. Third, integrate velocity changes into the output gate and dynamically reweight the output channel via vehicle velocity scaling. Vehicle motion changes the scale of the predicted pedestrian bounding box. As the vehicle approaches the pedestrian, the bounding box grows. The rescaling process captures the relative motion and updates the pedestrian's bounding box size accordingly. [1]

The authors T. T. Feng et al proposed a pedestrian detection algorithm based on attention mechanism and feature enhancement with SSD is presented because of low resolution of small targets, limited target features, noise interference etc which affect accuracy of detection. Channel feature fusion is used to fuse non-adjacent convolutional layers to gain significant edge gradient features and semantic information features. At last, by optimization of the attention mechanism CBAM, the channel features and spatial features are coupled under different fusion detection layers to improve the feature weight of the pedestrian's salient region. method can effectively detect the small target pedestrian in the image, and reduce the error detection and missing detection. [2]

K. Saleh et al has addressed the problem of intent action prediction of pedestrians. Paper describes the issue of understanding the behaviour and intention of humans are the main challenges faced by autonomous vehicles. Especially, is areas of city traffic and congested environment, detecting the intentions and actions of pedestrians become even harder. To address this problem image sequences from a monocular RGB camera are used and a real-time framework is proposed that can accurately detect, track and predict pedestrians' behaviour based on a tracking-by-detection technique in conjunction with a novel spatio-temporal DenseNet model. The model is trained and evaluated on framework based on real data collected from city traffic environments. [3]

The author S. Zhang et al proposed, pose estimation (key point detection) is used to generate pedestrians' variables from CCTV videos. The Problem of crosswalks at red-light crossing is discussed in this paper as it is a threat to traffic safety. few studies have used trajectory data in time sequence in pedestrian's red-light crossing. Four machine learning models are used to predict pedestrian intent when crossing a red light. it is found that key point variables such as the angle between ankle and knee (left) and elbow and shoulder (right) are shown to be important. This model can be implemented to prevent red light crosswalk accidents by alerting drivers. [4]

The author D. O. Pop et al has primarily focused on assessing whether pedestrian behaviour constitutes a dangerous situation time to cross the street. 1) proposes a pedestrian detection and action detection component based on RetinaNetv 2) Estimating the time it takes for multiple pedestrians to cross the street using recurrent neural network. For each pedestrian, the recurrent network estimates the pedestrian's intent. Predict when to cross the road. We will perform an experiment based on the JAAD dataset to demonstrate this integration. Multiple pedestrian action tags in detection part when merging with recurrent neural network (LSTM) Allows for significant performance improvements. [5]

The author Z. Yi et al has proposed pedestrian detection model based on tripartite fusion (PDMTF) is proposed based on the histogram of oriented gradient (HOG) and support vector machine (SVM) pedestrian detection algorithms, incorporating median filter (MF), The model first removes noise using a median filter algorithm Effectively reduce the effects of noise on images, HOG feature descriptor. Then, using the extracted pedestrian features, SVM classifier. To optimize the SVM classifier, The model performs secondary training on misidentified individuals Pedestrian only area. The final experimental result is The PDMTF model has a pedestrian false detection rate of only 7percent. High complex pedestrian detection rate neighbourhood environment. [6]

The author Zhang et al proposes a pedestrian detection acceleration algorithm called non-pedestrian area estimation (NPAE) Based on the concept of non-pedestrian area (NPA). The NPAE algorithm estimates and removes nonpedestrian regions in the image, followed by pedestrian detection in the NPAE source image. test the proposed NPAE algorithm on both GPU and CPU platforms using two datasets containing images with three different resolutions. RetinaNet is chosen as the reference pedestrian detector. [7]

The author C. B. Murthy et al proposed a robust Enhanced Tiny-Yolov3 Network is developed by introducing an antiresidual module .it helps to improve network's ability to extract features. Second, the loss function is improving and reduces bounding box loss error and optimize the network. Third prediction scale of size 26 x 26 is removed from the Tiny-Yolov3 network so the computational complexity of the network is reduced. The proposed network is trained on the extracted pedestrian images extracted from Pascal Voc-2007 dataset. Experimental results show this network improves the recognition accuracy while detecting smaller pedestrians and it still meets the real-time requirements. [8]

III. System Architecture

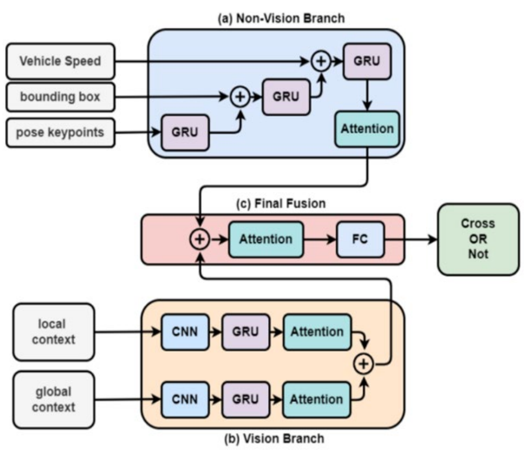

Fig.1. System Architecture [9]

Three sections make up the architecture: the vision branch, the non-vision branch, and the final fusion branch. The vision branch includes the CNN, RNN, and attention modules that combine various capabilities. Through dataset, local and global context are derived. To create the CNN module, we pre-train the VGG19 model on the ImageNet dataset. We create the RNN module using a gated recurrent unit (GRU). Because a GRU is more computationally effective than an LSTM, it should be used instead. Convolutional neural networks have an attention module for spatial attention. CNN is taught and focused more on the vital information using attention modules than on unhelpful previous knowledge. The vision branch combines two visual aspects, including global context (semantic segmentation of significant objects over the entire image) and local context (enlarged pedestrian appearance around the bounding box). By first collecting spatial characteristics from the CNN module and then extracting temporal features from the GRU module, local context is encoded. The same method is used to encode global context. The final visual feature is created by concatenating both local and global features after they have been supplied into their attention modules.

Three non-visual elements are combined in the non-vision branch (bounding boxes, pose key points, and vehicle speed). In accordance with their degree of abstraction and complexity, they are fused hierarchically. The impact of the fused features on the final prediction will increase as the fusion step progresses. The merging of visual elements is indicated by the yellow portion. Local context and global context 2D convolutional features are encoded by GRUs and fed to the attention blocks separately. Final visual features are created by concatenating the two outputs. The blue portion indicates the merging of regional elements (non-visual). To create the final non-visual features, these non-visual features are encoded by a different GRU, combined hierarchically, and then given to an attention block. The complete fusion is seen by the red portion. An attention block receives the concatenated final visual and final non-visual elements. The final prediction is then made using a fully-connected (FC) layer. [9].

IV. Conclusion

The proposed deep learning model is capable of predicting whether person is going to cross the road or not. In this model, CNN and YOLO algorithm is used to tackle the problem of prediction of the pedestrian intension. It contains modules as data acquisition, data pre-processing and CNN and YOLO model with desktop application. In this firstly, video as taken as input and pre-processing is done on video. Later on, CNN and YOLO model is built. Lastly, model is applied to the region of the interest. Model had accurately classified pedestrian intension.

References

[1]. R. Quan, L. Zhu, Y. Wu and Y. Yang, "Holistic LSTM for Pedestrian Trajectory Prediction", in IEEE Transactions on Image Processing, vol. 30, pp. 3229-3239, 2021.

[2]. T. T. Feng and H. Y. Ge, "Pedestrian detection based on attention mechanism and feature enhancement with SSD", 2020 5th International Conference on Communication, Image and Signal Processing (CCISP), 2020, pp. 145-148.

[3]. K. Saleh, M. Hossny and S. Nahavandi, "Real-time Intent Prediction of Pedestrians for Autonomous Ground Vehicles via Spatio-Temporal DenseNet", 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 9704-9710.

[4]. S. Zhang, M. Abdel-Aty, Y. Wu and O. Zheng, "Pedestrian Crossing Intention Prediction at Red-Light Using Pose

Estimation", in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 2331-2339, March 2022.

[5]. D. O. Pop, A. Rogozan, C. Chatelain, F. Nashashibi and A. Bensrhair, "Multi-Task Deep Learning for Pedestrian Detection, Action 42 REFERENCES Recognition and Time to Cross Prediction", in IEEE Access, vol. 7, pp. 149318-149327, 2019.

[6]. Z. Yi and H. Xiaoyong, "Research on Pedestrian Detection System based on Tripartite Fusion of HOG+SVM+Median filter", 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2020, pp. 484-488.

[7]. H. Zhang, D. O. Pop, A. Rogozan and A. Bensrhair, "Accelerate High Resolution Image Pedestrian Detection with Non-Pedestrian Area Estimation", in IEEE Access, vol. 9, pp. 8625-8636, 2021.

[8]. C. B. Murthy and M. Farukh Hashmi, "Real Time Pedestrian Detection Using Robust Enhanced Tiny-YOLOv3", 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1-5.

[9]. D. Yang, H. Zhang, E. Yurtsever, K. A. Redmill and U. Ozguner, "Predicting Pedestrian Crossing Intention with Feature Fusion and Spatio Temporal Attention", in IEEE Transactions on Intelligent Vehicles, vol. 7, no. 2, pp. 221-230, June 2022.