REDUCING THE RATE OF ACCIDENTS INVOLVING PEDESTRIANS IN TRAFFIC USING DL

Fatih Ay, Bayram Kaya & Fatma Akçakır
Department of Computer Science
Hacettepe University
Ankara, Türkiye
b21945815@cs.hacettepe.edu.tr

ABSTRACT

Pedestrian intention to cross is a topic that has long been included in active research. But since there is no common training and evaluation procedures, it is difficult to measure the progress towards the target. In this study, we will analyze the results using 2 different procedures.

1 INTRODUCTION

When pedestrians do not pay attention to cars while crossing the street, there is a high risk of pedestrian accident.
Our aim is to predict the time when pedestrians want to cross and to warn drivers in advance if it is a dangerous situation and by doing this we are aiming to reduce the number of accidents.
We need a model that understands the pedestrian's intention to understand whether the environment is dangerous when driving a car. It's not that hard to find the database containing this information. The real challenge is pedestrian's intention estimate. To find out people's intentions, you have to either do a live poll or guess yourself.
The theme you chose "Computer vision for emergencies". In the paper you gave for this theme, there is a section about "Emergencies caused by humans". We think the subject we are working on fits this category. We are trying to prevent accidents caused by humans(pedestrians and drivers)

2 RELATED WORKS

There are a lot of researches for preventing accidents caused by drivers (Civik & Yuzgec (2023), Iftikhar et al. (2022)). But our aim is to use intention of pedestrians.
There is various approaches for detect the intention of pedestrians, such as:

2.1 PEDESTRIAN DETECTION AND TRACKING

This is one of the basic approaches. With this approach they aim to predict pedestrian intent based on intrinsic pedestrian features such as poses. (Rehder et al. (2017) , Varytimidis et al. (2018))

2.2 TRAJECTORY PREDICTION

In this approach, they aim to predict where the pedestrian will be in the future and use this information to infer the pedestrian's intention. (Habibi et al. (2018))

2.3 ACTION PREDICTION

This approach predicts pedestrian intent by using the relationship between the past, current and potential future information(Farha et al. (2018))

2.4 Scene graph parsing and visual reasoning

This approach predicts pedestrian intent by using the Spatio-temporal relationship between the various objects in the scene (Zellers et al. (2017))

3 Method

We are planning to test SF-RNN(Rasouli et al. (2020)) model and MultiRNN(Bhattacharyya et al. (2017)). Recurrent neural networks (RNNs) are a particular kind of neural network where the results of one step are fed into the next stage's input.
We are planning to make these models using Gated recurrent unit. A Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN). GRU can process sequential data such as text and speech. The fundamental principle of GRU is to selectively update the network's hidden state at each time step via gating techniques. The reset gate and the update gate are two of the GRU's two gating systems. The update gate determines how much of the new input should be used to update the hidden state, while the reset gate determines how much of the prior hidden state should be forgotten. (Gupta (2023))
If you want to know more about the GRU you can check out the source (Kostadinov (2017))

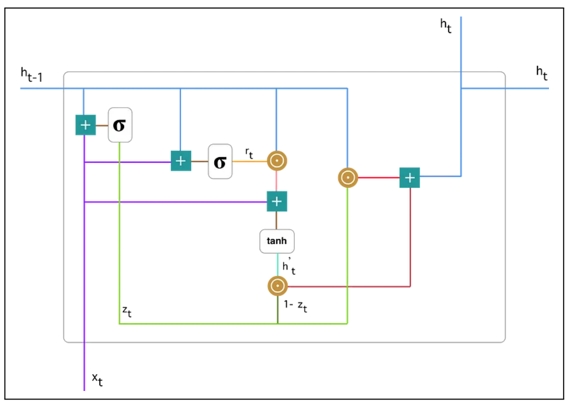

Figure 1: A gated recurrent unit

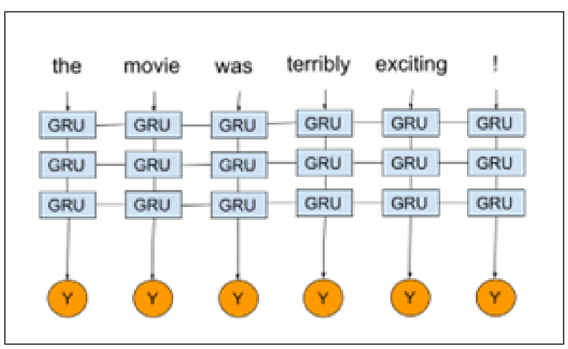

Figure 2: Multiple RNNs

Multi-layer RNNs are the applying multiple RNNs on top of each other, as seen in this image(Anonymous (2019))

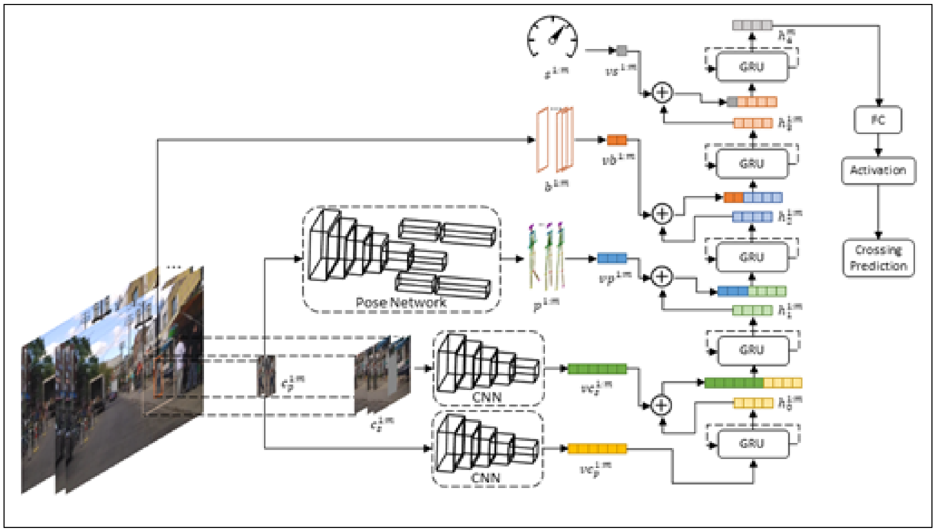

Figure 3: SF-RNN

SF-RNN model is based on this architecture. This is a stacked RNN architecture in which we gradually fuse the features at each level according to their complexity

4 EXPERIMENTAL SETTINGS

We are using dataset that named PIE(Rasouli et al. (2019)). This dataset was made to examine pedestrian behavior in traffic. They recorded 6 hours of HD video with on-board camera at 30 FPS and split into approximately 10 minute chunks. The dataset contains accurate ego-vehicle information from OBD sensor such as speed, spatial annotations such as traffic lights, behavioral annotations such as walking and intention annotations. PIE dataset is one of the largest publicly accessible datasets for analyzing pedestrian traffic, the collection includes over 290K annotated video frames with 1842 pedestrian samples. The PIE dataset also includes 429 pedestrians who do not intend to cross, 519 pedestrians who do mean to cross but do so in front of the vehicle, and 894 pedestrians who intend to but do not cross.
There is 6 different sets in the PIE dataset. We plan to use set 02, set 03, and set 06 to train, set 04 and set 05 for validation, and set 03 to test. Due to our hardware constraints we are using 1 video from each set for now. Here, you can find some extra statistics about the PIE dataset(Note that, this statistic is about all data):

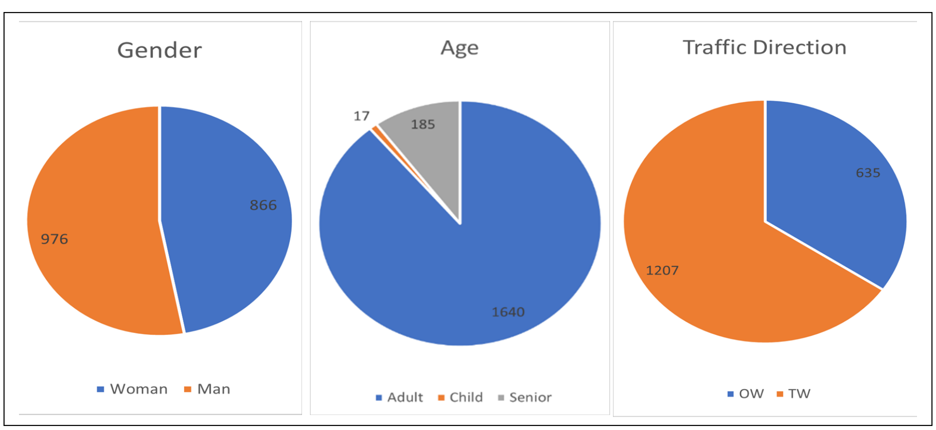

Figure 4: Summary of the dataset

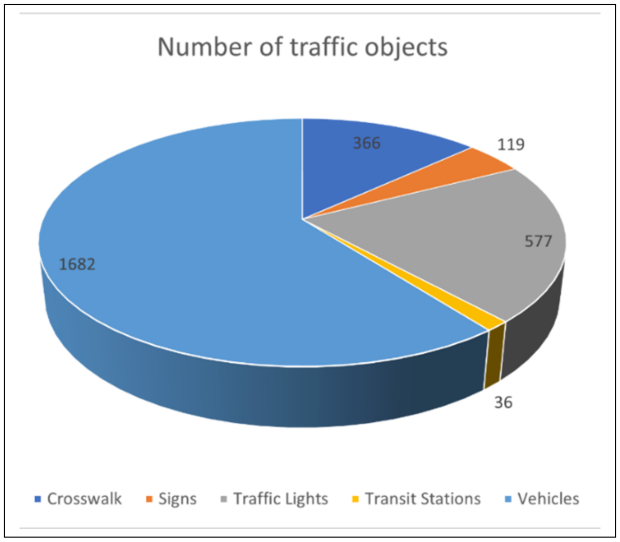

Figure 5: Summary of the dataset

Each pedestrian's observation data is sampled so that the last frame was between 1 and 2 seconds (or 30–60 frames) before the crossing event began.
All models have a set observation length of 16 frames.
Additionally, we plan to use optical flow data from FlowNet2 and pose data from OpenPose (18 body joint coordinates concatenated into a 36D feature vector) if needed.
We are considering testing our method for 2 different architecture with 2 different learning rates and we are planning to use F1 score to evaluate the performance of our models. As a result of F1 score comparisons, we aim for our best model to have an F1 score above 45 percent. Our method is similar to this study (Kotseruba et al. (2021)) Taking into account the results achieved in this study and the limitations of the equipment we have, we reached the figure of 45 percent. By hardware constraints we mean that we are using only a small portion of the dataset.
I didn't mention what we are going to test. In order to anticipate whether a pedestrian would begin crossing the street at some time t given an observation of length m, we construct the pedestrian action prediction as a binary classification issue.

5 Initial results

There is no initial result.

References

Esra Civik and Ugur Yuzgec. Real-time driver fatigue detection system with deep learning on a low-cost embedded system. 2023.

Sundas Iftikhar, Zuping Zhang, Muhammad Asim, Ammar Muthanna, Andrey Koucheryavy, and Ahmed A. Abd El-Latif. Deep learning-based pedestrian detection in autonomous vehicles: Substantial issues and challenges. 2022.

Eike Rehder, Florian Wirth, Martin Lauer, and Christoph Stiller. Pedestrian prediction by planning using deep neural networks. 2017.

Dimitrios Varytimidis, Fernando Alonso-Fernandez, Boris Duran, and Cristofer Englund. Action and intention recognition of pedestrians in urban traffic. 2018.

Golnaz Habibi, Nikita Jaipura, and Jonathan P. How. Context-aware pedestrian motion prediction in urban intersections. 2018.

Abu Farha, Alexander Richard, and Juergen Gall. When will you do what? - anticipating temporal occurrences of activities. 2018.

Rowan Zellers, Mark Yatskar, Sam Thomson, and Yejin Choi. Neural motifs: Scene graph parsing with global context. 2017.

Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. Pedestrian action anticipation using contextual feature fusion in stacked rnns. 2020.

Apratim Bhattacharyya, Mario Fritz, and Bernt Schiele. Long-term on-board prediction of people in traffic scenes under uncertainty. 2017.

Alind Gupta. Gated recurrent unit networks. 2023.

Simeon Kostadinov. Understanding gru networks. 2017.

Anonymous. Multi-layer recurrent neural networks in keras. 2019.

Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John K. Tsotsos. Pedestrian intention estimation (pie) dataset. 2019.

Iuliia Kotseruba, Amir Rasouli, and John K. Tsotsos. Benchmark for evaluating pedestrian action prediction. 2021.