IOSUD - Politehnica University of Timisoara
Doctoral School of Engineering Studies

Using deep neural networks in predicting the movement of road users

Ph.D. thesis – Abstract
To obtain the scientific title of doctor
Politehnica University of Timișoara
In the doctoral field Electronic Engineering, Telecommunications, and Information Technology

Author Eng. Bogdan Ilie SIGHENCEA
scientific coordinator Prof. Univ. Dr. Habil. Ing. Cătălin Daniel CĂLEANU
September 2023

The doctoral thesis is structured into seven chapters. Chapter 1 presents a brief introduction to the topic of using deep neural networks to predict the trajectory of pedestrians, motivating the need for such systems. Chapter 2 deals with elements of artificial intelligence with a focus on the concept of artificial neural networks. Concepts related to the architecture and training modality related to deep neural networks are presented, especially models used in the problem of pedestrian trajectory prediction (PTP). The analysis of the state-of-the-art achievements of PTP is carried out throughout Chapter 3. Successful resolution of the PTP problem depends on the sensory information available to the prediction system. For this reason, it was chosen to analyse the characteristics and performance of various types of sensors in Chapter 4 of the thesis. The proposed solutions shall be presented together with the related experimental results in Chapters 5 and 6 respectively. The paper ends with a chapter devoted to conclusions and possible directions for further development. The paper also contains a section on bibliographic references.

In the figure below you can trace the entire structure of the work divided into chapters and subchapters.

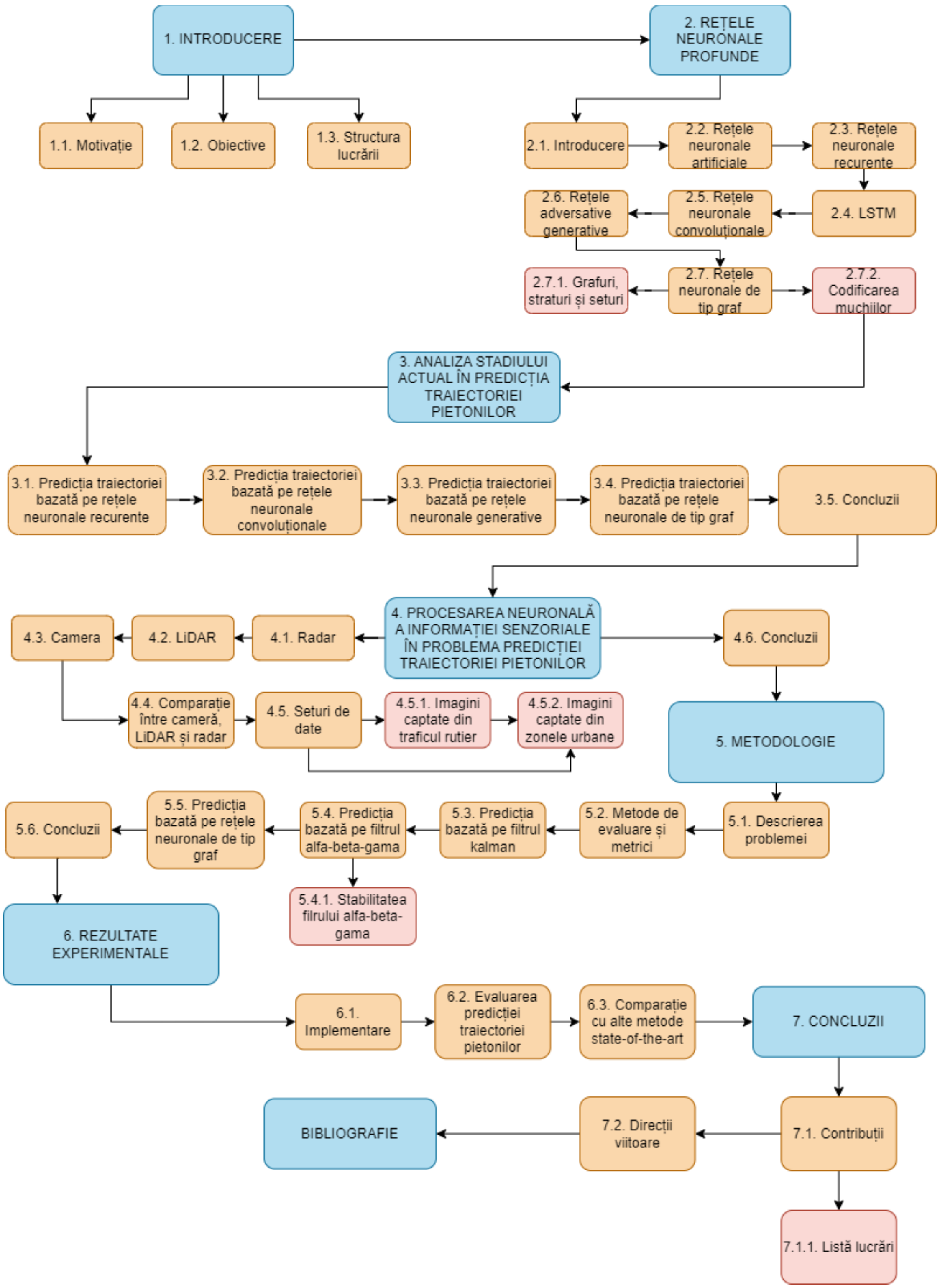

Figure 1. The structure of the paper by subchapters.

1. Introduction

In recent decades, automakers have been constantly working to improve the driving experience and make road vehicles safer by developing driver assistance technologies. To assess the extent of advances in driver assistance technology, the Society of Automotive Engineers (SAE) defined six levels of autonomy. These levels range from 0, which corresponds to fully manual driving, to 5 fully autonomous, which is the goal of recent research conducted by both the automotive industry and the academic world.

Autonomous vehicles (VAs) have the potential to transform the world as we know it, revolutionising transportation, making it faster, safer, and less labour-intensive. It is important for AV systems to accurately perceive and react safely to various real-world driving scenarios. This requires AV perception systems to understand the behaviour of surrounding road users (e.g. vehicles, pedestrians, and cyclists) and accurately predict their future trajectories and behaviours [1]. For short-term predictions, it may be acceptable to use approaches based on pure physics. However, since future scenarios are unknown, a long-term prediction system is essential to allow not only modelling interaction between different agents, but also to identify traversable regions defined by road routes and traffic compliance.

According to the latest report on road accidents published by the European Road Safety Observatory [2], almost 24,000 people died in EU road accidents in 2019, 4668 of whom are pedestrians. Of this total, 729 pedestrians who died in road accidents are from Romania. In terms of global statistics, the data is more worrying. The Global Road Safety Report [3], published by the World Health Organisation (WHO), indicates that more than 1.3 million people died in road accidents worldwide in 2017. About 23% of these deaths are pedestrians. On the other hand, in countries under the Organisation for Economic Co-operation and Development (OECD), more than 20 000 pedestrians lose their lives every year. Pedestrian deaths account for between 8% and 37% of all traffic fatalities, depending on the country and year [4]. Most of these tragic accidents occur in crowded areas at pedestrian crossings, with poor visibility due to low driver attention and/or fatigue.

2. Motivation

According to [5], the number of accidents with elderly pedestrians is influenced by multiple factors in the environment of the localities. As a result, reducing (or eliminating) these collisions is an important safety concern. In these situations, helping the driver includes predicting pedestrian behaviour. This helps reduce the effect of various factors that could negatively affect traffic safety (such as fatigue, poor visibility, accidental cognitive distraction, etc.).

In the event of an impact, pedestrians have virtually no protection. Therefore, reducing (eliminating) these impacts is an important safety issue. Helping the driver in such conditions includes predicting the trajectory and / or behaviour of the pedestrian and mitigating consistent driver errors (e.g., fatigue, cognitive thinking) and includes developing new technologies to reduce the number of accidents (by up to 93.5%, according to [6]).

We humans make important intuitive decisions based on the sequence of actions and interactions with other people in the scene to achieve safe and smooth navigation. This intuition allows movements that are very dynamic, since we can decide which route to take in a very dynamic manner. This simple but valuable information is crucial to deciding the next step. However, with the advent of deep learning, advanced algorithms are being developed that read pedestrian instincts and allow action. Different methods are explored, from trajectory prediction to behavioural analysis [7]. In this doctoral thesis, we investigate the idea of using RGB monocular images as basic information to predict the trajectory of pedestrians.

3. Objectives

This thesis aims to predict the trajectory of pedestrians in different scenarios using deep neural networks. This field is very successful but also highly studied by researchers due to the development of optical sensors (e.g., RGB camera, Radar, LiDAR, etc.) but also the emergence of new deep learning architectures. To achieve the proposed objective, several work tasks have been defined:

- Analysis of the current state in the field of pedestrian trajectory prediction, as well as analysis of all existing architectures using deep neural networks.
- Identify and implement new deep learning (hybrid) solutions that bring substantial improvements to existing solutions.
- Applying the developed models to the most well-known databases for this field and surpassing state-of-the-art solutions in terms of performance and accuracy.
- Testing the feasibility and limits of proposed methods in an extensive way under ideal conditions, using real-world databases.
- Measuring the influence of modelling four different pedestrian dynamics, i.e., standstill, starting, stopping, and walking. These dynamics allow pedestrian changes to be properly defined in real-life scenarios.
- Development of a method for predicting pedestrian routes, applying new neural models (e.g., CNN, LSTM and GNN).

4. Summary of thesis by chapters

4.1. Deep neural networks

To solve the problem of PTP, in recent years, several methods based on deep learning in the literature have been proposed. This section details the most used methods in this area, classified according to the type of DNN architecture. The identified pedestrian trajectory prediction methods based on deep learning used mostly three architectural structures, as follows.

With lower computing costs and faster communication that provides unlimited access to information and a better understanding of the physical world around us, highly automated decision-making, such as artificial intelligence, is becoming the engine technology of the twenty-first century. However, artificial intelligence has been the mainstay of many applications, such as autonomous cars, digital assistants, and medical imaging, just to name a few. However, could there be a lack of proper understanding of this critical technology? Also, due to the huge hype surrounding this technology, there are many misunderstandings in terminology. This misinterpretation can be observed mainly when the terms artificial intelligence, machine learning, and deep learning are constantly changed. Although the terms seem equivalent, the meaning of each term varies, and this section aims to clearly articulate the differences between artificial intelligence, machine learning, and deep learning.

Igor Aizenberg and his colleagues first coined the term DL in 2000 [8]. Deep learning algorithms are a subcategory of machine learning algorithms that mimic the human learning process through machine learning through examples. Deep learning algorithms use complex multilayered learning structures known as neural networks, which learn an implicit representation of raw data autonomously to produce the desired result. In other words, to make traditional machine learning algorithms work, an essential but extremely complicated step known as feature extraction is required, which must be done manually by experts in the field for which algorithms to work. On the other hand, deep learning algorithms learn these automatically extracted characteristics, while the learning structures in these algorithms optimize to obtain the best possible abstract representation of input data.

Different trajectory models of motion (with multiple origins and destinations) and dynamic human interaction are key to a trajectory prediction model under complex circumstances. Most of the existing methods based on deep learning are heavily dependent on specific scenarios because they perform trajectory prediction using absolute coordinates. The trajectory of the movement is a relative movement that coincides with time, and human interaction is a relative movement between pedestrians. This motivates the construction of a trajectory prediction model for the relative motion of both the trajectory of motion and human interaction.

Different deep learning architectures have been designed to address specific tasks in different areas. For example, recurrent RNA and attention mechanisms have been used primarily for language modelling in natural language processing. And convolutional networks are used extensively to solve the problem of image classification and object recognition in the field of artificial vision. These architectures have also been expanded and adapted to other areas, such as financial forecasting and climate understanding. They are also used in the field of autonomous driving and specifically for motion forecasting.

4.2. Analysis of the status in pedestrian trajectory prediction

This chapter presents the stage of evolution of studies and working methods related to the problem of predicting the trajectory of pedestrians. We proposed a taxonomy of methods for better understanding and classification of different approaches to the problem [9].

To solve the PTP problem, several methods based on deep learning have been introduced into the literature in recent years. This chapter details the most used methods in the field, classified by DNN architecture type. The identified methods for predicting the trajectory of the pedestrians, based on methods of in-depth study, mainly used four architectural structures. These will be described below. A correspondence between the neural network architectures used in the literature and the PTP problem can be seen in Figure 2.

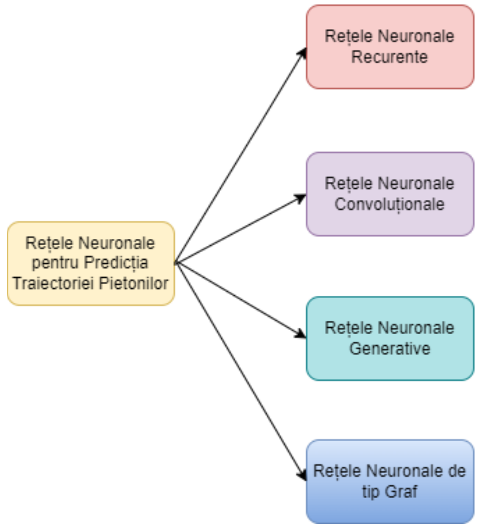

Figure 2. The most widely used deep neural network to predict the trajectory of pedestrians.

Architectures used to predict pedestrian trajectory include recurrent, convolutional, generative, and graph neural networks. The estimation of the trajectory with them is presented in the following paragraphs.

4.3. Neural processing of sensory information in the problem of predicting pedestrian trajectory

This chapter examines representative works in the field ("State-of-the-art", SOTA), especially prediction methods that use deep neural networks together with the most used sensors (camera, LiDAR, and radar) for autonomous driving. The performance of each car sensor (radar, LiDAR, and camera) was presented comparatively by highlighting their advantages and disadvantages in different tasks, e.g., object detection and classification, distance estimation, detection of road markings and road boundaries, etc. The most important publicly available data sets used by researchers in the implementation of PTP solutions have been identified.

An autonomous vehicle (VA) is an assembly equipped with sensors (LiDAR, radars, cameras, etc.) and systems capable of automatically controlling it so that it can drive on the road without human intervention. To accomplish such tasks, the control system must be able to detect objects near the vehicle and estimate their future trajectories and the kinematic parameters of their motion [10].

Sensors and systems that process incoming information assist drivers by signaling different circumstances to minimize the risk of exposure. They can even automate driving tasks to eliminate human error [11]. To collect information from the outside environment, VA uses sensors called exteroceptive. The processing of the information received leads to the recognition of other road users (pedestrians, vehicles, etc.) and nearby objects. In recent years, external VA sensors have gained importance, especially due to the development of image processing and cameras, as these systems enable a wide range of applications [12].

According to [13], autonomous driving assistance is mainly based on systems related to the processing of images from cameras. LiDAR is the most necessary sensor in automotive systems. Unlike cameras, it is characterized by omnidirectional detection and is not affected by light conditions. A quarter of the automotive sensor market is ultrasonic and radar sensors, while other exteroceptive sensors, such as microphones, account for 18% of the market (see Figure 3).

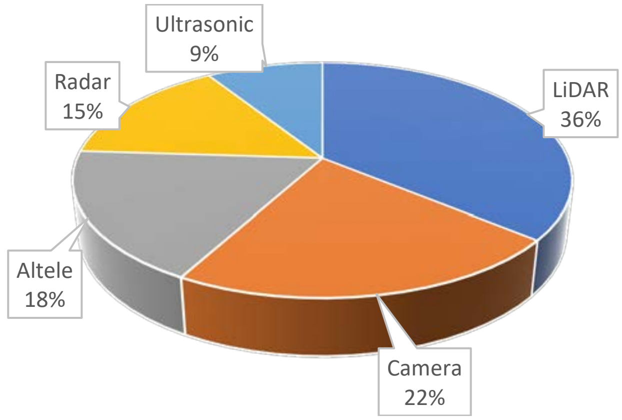

Figure 3. Prediction of automotive sensor market growth (compound annual growth rate (CAGR), 2017-2022) on exteroceptive sensors, according to [79].

A set of sensors placed on vehicles captures data that is then processed to obtain a digital representation of the environment. This information is used for risk assessment, as well as planning and predicting trajectories, which leads to vehicle movement control.

Perception provides how the vehicle can obtain information about what is happening in its operating environment. In an autonomous driving scenario, there are numerous road users around the vehicle: pedestrians, cyclists, motorcyclists, other vehicles, etc. A major challenge is to perceive all these elements continuously and accurately, without having false alarms (reporting non-existent obstacles) or lack of detection (i.e., omission of real obstacles). For this purpose, several active and passive exteroceptive sensors can be used, including different types of cameras, LiDAR, and RADAR. State-of-the-art detection algorithms also provide classification of perceived objects.

However, there is no single sensor technology capable of providing accurate and complete space-time information about everything around the vehicle, each with its own advantages and disadvantages [14]. The solutions applied today thus combine different types of sensors by means of a process of interconnection of several sensors (fusion).

To test pedestrian trajectory prediction systems, researchers typically use multiple datasets. They provide images of pedestrians in different scenarios (promenade, pedestrian crossings, sidewalks, etc.). In these people move in different directions.

Large datasets are important resources for training and testing DNN models. For this, the data of these are labelled. In a possible scenario, the pedestrian can be instructed to perform predefined actions (stopping, crossing on crosswalks, crossing, etc.), but is not limited to them. Unfortunately, the amount of data collected is limited. Two major challenges can be identified in the available datasets: because they are predefined, they do not encompass all the information (the registered pedestrian is an 'actor', the variability of his actions is non-existent). A second challenge is represented by the fact that in real life, pedestrians' actions are variable, they are determined by random events (the arrival of the bus, the traffic light changes color, etc.). Real-life scenarios were used for low-level models such as detection and tracking. Unfortunately, they do not provide the necessary data for higher-level models (e.g., social interactions).

4.4. Methodologies

In this chapter, the methods used throughout the research to predict the trajectory of pedestrians were described. In connection with the theme of this chapter we have published papers [15], [16] for prediction based on the alpha-beta-gamma filter. An analogy was drawn between the Kalman filter and the alpha-beta-gamma filter to identify the accuracy of the method. For graph-based neural network prediction, we published the papers [17], [18]. Here were presented the methods and architecture of prediction solutions using graph-based spacetime convolution neural network (ST-GCNN) and temporal extrapolation convolution neural network (TXP-CNN). For the evaluation of the methods and their comparison with other methods in the literature, the most representative metrics in the field were presented.

The main objective of trajectory prediction is to estimate the future positions of a group of *N* pedestrians in a real scene. This is based on their previous and current positions on a map representation over a time interval, starting from the moment T_o (as seen in Figure 4) after iterations over time. The actual position of each pedestrian in a scene is represented by the real coordinate X = (x, y) at iteration $T_p i$, which is denoted by $X_t^i = (x_t^i, y_t^i)$. Here $t \in \{1, \ldots, T_o\}(x_t^i, y_t^i)$are random variables that describe the probability distribution of the pedestrian position *n* at time *t* within the scene (*n* is scene dependent). The actual/true future trajectory is denoted , where .$Traj_{obs} = \{Y_t^i = (x_t^i, y_t^i)\} i \in \{1, \ldots, n\},\ T_o + 1 \leq t \leq T_p$

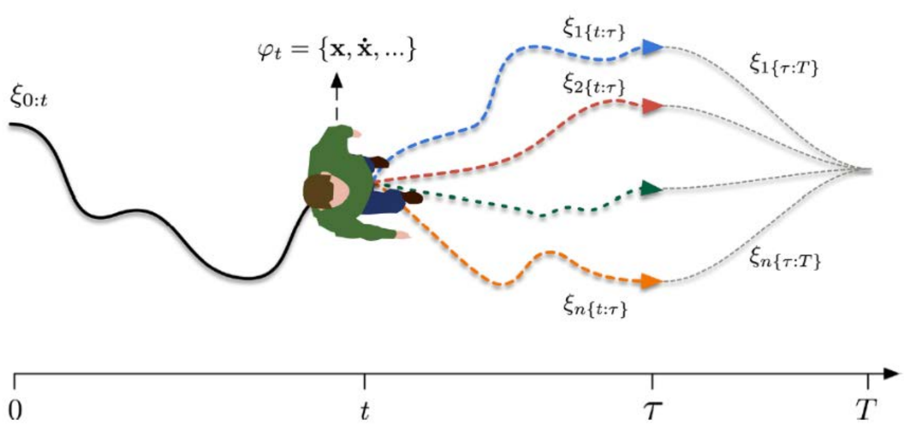

Figure 4. The temporal distribution of each pedestrian's position from moment T_o to time T_p. Three types of positions can be mentioned here: observed, real future, and predicted [17].

The predicted positions represent a series of random variables. They are the result of an assumption that the position of pedestrian $Traj_{pred} = \{\hat{Y}_t^i = (\hat{x}_t^i, \hat{y}_t^i)\} i \in \{1, \ldots, n\},\ T_o + 1 \leq t \leq T_p i$ at time t follows a two-dimensional Gaussian distribution, denoted . Within this distribution it represents the center of the group of pedestrians at time $\hat{Y}_t^i \sim \mathcal{N}(\mu_t^i, \sigma_t^i, \rho_t^i) \mu_t^i = (\mu_x, \mu_y)_t^i t$. The standard deviation of the distribution and the correlation coefficient are denoted.

To obtain trajectory prediction, the proposed architecture predicts the parameters of the Gaussian distribution.$\sigma_t^i = (\sigma_x, \sigma_y)_t^i \rho_t^i (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)_t^i$.

To estimate positions at times for each road user, the locations observed at time points are used. A neural network architecture is used to estimate the trajectory. $T_o + 1 \leq t \leq T_p 1 \leq t \leq T_o$. To train the model parameters, the function of loss of the negative logarithm of probability is used, as shown in Equation (1).

$$L^i(W) = -\sum_{t=1}^{T_p} log(f(x_t^i, y_t^i | \mu_x, \mu_y, \sigma_x, \sigma_y, \rho)) \tag{1}$$

where W represents the parameters of the driven network. The amount of loss is minimized to achieve optimal network performance.

4.4.1. Assessment Methods

In artificial vision research, trajectories are often described by motion statistics such as the number of collisions, average acceleration, average speed, and total distance travelled [19]. For each road user, there are eight observation stages (3.2 seconds). For all datasets, twelve steps (4.8 seconds) are used to represent actual future positions. Euclidean L2 norms shall be used to assess the difference in position between the actual and estimated trajectory.

ADE and FDE are the two main metrics used to evaluate deterministic regressions. Although these metrics are natural to the intended goal, easy to implement, and interpretable.

Figure 5 (a) illustrates one of the most common ways to directly compare adjacent trajectories, i.e. measure how far apart it is for each t and then averse these distances to get the mean error over the duration of the forecast. This is commonly known as average displacement error (ADE) and is usually reported in units of length, e.g., meters:

$$ADE = \frac{\sum_{n=0}^{N} \sum_{t=0}^{T_p} \| \hat{p}_t^n - p_t^n \|_2}{N \times T_p} \tag{2}$$

Often, it is possible that the interest is represented by the error of the end point of the subject's trajectory, illustrated in Figure 5 (b) (in particular, only the points). This provides a measure of the method error at the end of the prediction horizon and is commonly referred to as the Final Displacement Error (FDE). And it is usually reported in units of length$\hat{p}_3 p_3$:

$$FDE = \frac{\sum_{n=0}^{N} \| \hat{p}_t^n - p_t^n \|_2}{N}, t = T_p, \tag{3}$$

where N represents the number of pedestrians, the number of predicted time steps, is the actual result and the predicted result at time step $T_p p_t^n \hat{p}_t^n t$.

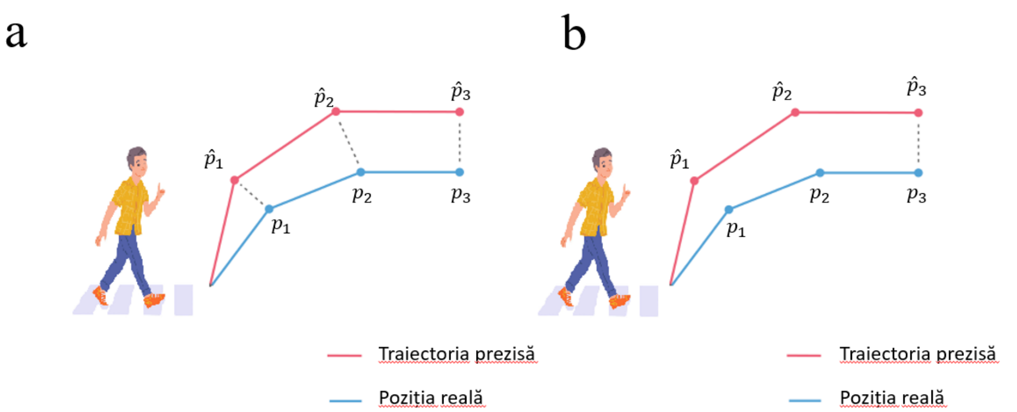

Figure 5. Illustrations of metrics. (a) Mean displacement error (ADE), (b) Final displacement error (FDE).

4.4.2. Prediction based on graph-like neural networks.

This subsection provides a description of the key aspects of the proposed estimation method. Subsequently, the design of each module will be discussed in detail, including the general creation of the model. The main goal is to predict future trajectories for numerous interacting factors using position history and context information. The prediction method is also extensible to multitarget tracking frameworks [17].

The method consists of two components: graph-based spacetime convolution neural network (ST-GCNN) and temporal extrapolation convolution neural lattice (TXP-CNN). The first component uses convolution to extract features. These characteristics provide a concise description of the history of observed trajectories. They are fed into TXP-CNN, which uses them to predict the positions of all pedestrians in the group and to extrapolate future trajectories. The overall diagram of this method is illustrated in Figure 6.

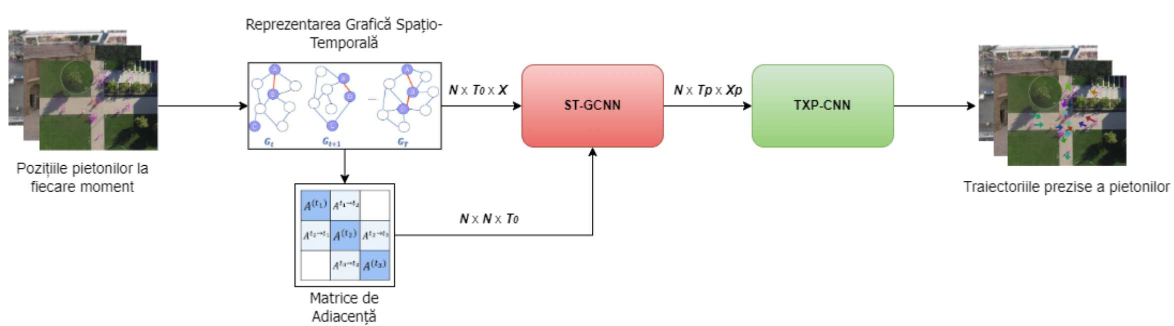

Figure 6. General architecture of the proposed method. Optimizing the layer size allows increased accuracy in trajectory prediction.

The first component of the architecture is ST-GCNN. It is used to incorporate the representation of the object [20]. It is designed to extract the space-time embedding from the input graph.

The second component used is TXP-CNN. This works directly on the temporal dimension of graphic embedding V that grows to meet accuracy requirements in prediction. The second component has fewer parameters than recurring units, since it is based on convolution operations in the feature space. It is important to note that the TXP-CNN layer is not invariant to permutation, since variations in graphics embedding lead to different results. However, predictions are invariant if the pedestrian order is altered [18].

The entire process of predicting the pedestrian trajectory is illustrated in Algorithm 1.

Algoritmul 1 Rețele Convoluționale cu Graf Dinamic
Date de intrare: coordonatele pietonilor $X = (x, y)$;
Date de ieșire: metricele de assessment, eroare medie de deplasare (ADE) și eroarea deplasării finale (FDE);

for $t \in [1, T_o]$ **do**
    reprezintă traiectoriile drept un graf: $G_t = (V_t, E_t)$;
    calculează traiectoriile viitoare; $Traj_{obs} = Y_t^i = \left(x_t^i, y_t^i\right)$
end for
creează distanță de distribuție N x N matrice de adiacență A_t^m utilizând ecuația (5.5.1);
generează matricea laplaciană;
for fiecare $i \in 1, \ \dots, \ n$ **do**
    for all $t \in [1, T_o]$ **do**
        distribuția probabilității pentru traiectoria prezisă: $\hat{Y}_t^i = (\hat{x}_t^i, \hat{y}_t^i) \sim \mathcal{N}(\mu_t^i, \sigma_t^i, \rho_t^i)$;
    end for
end for
colectează toate locațiile prezise si cele reale pentru fiecare pieton;
calculează ADE și FDE;
return ADE și FDE

Figure 7 shows the overall architecture of the optimized method. Specifically, the trajectory tree is first created to provide a discretely structured space. Similarly, the observed spatial interaction and trajectory are encoded one after another to obtain the interaction encoding and the observed encoding $\boldsymbol{P}_{coarse}$.

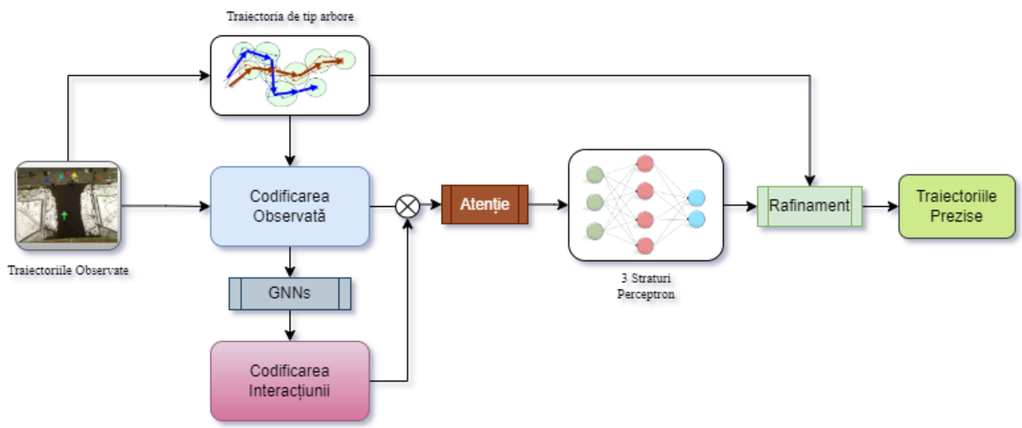

Figure 7. General architecture of the proposed method [21].

Another approach used in the paper [18] was to implement an encoder-decoder architecture, as can be seen in Figure 8. Called the Spatio-Temporal Graph Convolutional Network (ST-GCNN), it is used to model the temporal interactions of pedestrian trajectories observed over time. Here, a network of time graphs is used to model the temporal interactions of a single pedestrian in different time steps. The time graph of pedestrian *N* is created and represents its relative positions in different time samples.

The decoder contains another module, the "Time-Extrapolator Convolutional Neural Network" (TXP-CNN), which is used to predict next steps. TXP-CNN works directly on the temporal dimension of the chart incorporation and augments it as a necessity for prediction. TXP-CNN has fewer parameters than recurring units because it relies on convolution operations on entities. A special feature about the TXP-CNN layer is that it is not invariant to permutation since changes in chart embedding just before TXP-CNN lead to different results.

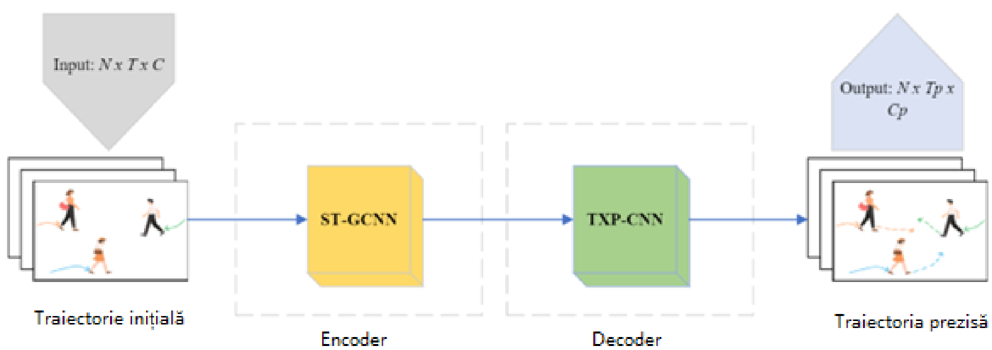

Figure 8. Encoder-decoder architecture

4.5. Experimental results

In this chapter, experiments and assessments performed in terms of pedestrian trajectory prediction are presented. As mentioned in Chapter 5, the experiments were conducted using pedestrian trajectory prediction methods based primarily on the use of deep neural networks. We evaluate the performance of our methods against similar trajectory prediction methods based on deep neural networks.

Trajectory prediction plays a role in estimating pedestrian risk potential, as it can also be used to predict crossing, which is necessary for collision prevention. To assess the applicability of trajectory prediction methods, we evaluate the performance of the methods we have proposed, quantitatively and qualitatively, using ETH [22], UCY [23], and Stanford Drone datasets [24], as they are widely used in the literature, are publicly available and use real-world coordinates. The use of data-driven methods requires the availability of sufficient quality data. Specifically, to predict pedestrian trajectory, the available data can be in two different formats: image coordinates or real-world coordinates. Image coordinates mean that each pedestrian is represented with the pixels they occupy in the camera image, while real-world coordinates mean that each pedestrian is represented by their position in meters, originating at an arbitrary point in the world. Qualitative assessments are performed by visualizing generated trajectories.

It will qualitatively analyse how the methods we proposed capture social interactions between pedestrians and take them into account when predicting distributions. We present cases where D-STGCN [17], PTPCNN [18] and TreeGNN [21] successfully predict collision-free trajectories between pedestrians from different angles, maintain parallel walking, and correctly predict the outcome of situations where a person encounters a group of pedestrians.

4.6. Conclusions

Pedestrian trajectory prediction is one of many areas where the advent of deep neural networks has completely changed the approach to the problem. Before 2016, models based on physical dynamics were the only way to accurately predict future pedestrian trajectories, but in 2016 it was demonstrated that deep learning could do this better. Since then, various new architectures have been proposed with increasingly better results. Data-driven models, such as models using convolutional, recurrent, or graph-like neural networks, are very sensitive to the quantity and quality of training data and how these data are then integrated into the model.

In this thesis, several aspects were analyzed, including data, sensors, and current state-of-the-art methods applied to the pedestrian trajectory prediction problem. Although the process of developing reliable solutions has been very laborious to date, it is still necessary to make a greater effort to achieve a system that ensures pedestrian safety on the streets. The selected articles were examined on several common factors to ensure a simple comparison for those interested in this topic, but also for the readers of this thesis.

By using deep learning approaches, current systems can better solve the PTP problem. These methods assume a series of pedestrian locations in the last few seconds and produce several future locations.

4.6.1. Contributions

In paper [17] we approached a new technique for predicting the trajectory of pedestrians using GNN-type neural networks. This solution introduces a two-component approach: a spatial graph neural network (SGNN) for interaction modelling and a temporal graph neural network (TGNN) for motion feature extraction. SGNN uses an attention method to periodically collect spatial interactions between all pedestrians. TGNN also uses an attention method to collect each pedestrian's pattern of temporal motion. With a smaller variable size (data and model) and a better prediction rate, D-STGCN is more compact and efficient than other SOTA solutions, providing better experimental results considering mean displacement error (ADE) and final displacement error (FDE) values, (Table 1 can be consulted).

Table 1. Quantitative results of state-of-the-art methods for ETH, UCY and SDD datasets defined in terms of ADE/FDE metrics. The AWG column represents the average results between scenes of datasets ETH-UCY

Methods	**SDD**	**ETH**	**HOTEL**	**UNIV**	**ZARA1**	**ZARA2**	**AWG**
Paper [17]	15.18/25.50	0.63/ 1.03	0.37/ 0.58	0.46/ 0.78	0.35/ 0.56	0.29/ 0.48	0.42/ 0.68
Paper [21]	15.84/25.17	0.64/ 1.07	2.29/ 1.85	0.41/ 0.67	0.27/ 0.43	0.24/ 0.38	0.77/ 0.88

Another solution we published [21] presents a tree based GNN approach to meet the challenge of multimodal prediction. The tree is designed based on observed data and is also used to predict future trajectories. Compared to previous approaches that use implicit latent variables to describe possible future trajectories, motion behaviors can be directly represented with a tree approach (e.g., walk straight and then turn left) and thus are offered by the model more socially appropriate trajectories.

To obtain information on the interaction between pedestrians, but also between pedestrians and the environment, we proposed a solution [18] based on two components (encoder and decoder). This solution was implemented using GNN+CNN and very good results were achieved in terms of ADE and FDE for PTP.

This article [25] analyzed the latest deep learning-based solutions for the pedestrian trajectory prediction problem along with the sensors used (camera, LiDAR, and radar) and related processing methodologies. The performance of each car sensor (radar, LiDAR, and camera) was presented comparatively by highlighting their advantages and disadvantages in different tasks, e.g., object detection and classification, distance estimation, detection of road markings and road boundaries, etc. The most important publicly available datasets used by researchers in implementing PTP solutions and performance indicators used in the evaluation process have been identified.

In the articles we published [15], [16] methods for prediction based on the alpha-beta-gamma filter were described. An analogy was drawn between the Kalman filter and the alpha-beta-gamma filter to identify the accuracy of the method. The results showed that the use of such architectures can provide a reliable solution for PTP.

Selected references

[1] S. Lefèvre et al., „A survey on motion prediction and risk assessment for intelligent vehicles”, Robomech J, 2014.

[2] European Road Safety Observatory, „Annual Accident Report”, 2020.

[3] WHO, „Global Status Report on Road Safety”, 2018.

[4] ITF, Pedestrian Safety, Urban Space and Health, 2012.

[5] D. Gálvez-Pérez et al., “The Influence of Built Environment Factors on Elderly Pedestrian Road Safety in Cities: The Experience of Madrid”, Int. J. Environ. Res. Public Health, Vol. 19, 2022.

[6] T. Winkle, „Safety benefits of automated vehicles: Extended findings from accident research for development, validation, and testing”, Autonomous Driving. Springer, 2016.

[7] S. Ahmed et al., „Pedestrian and cyclist detection and intent estimation for autonomous vehicles: A survey”, Applied Sciences, Vol. 9, 2019.

[8] I. N. Aizenberg, N. N. Aizenberg, and J. P. Vandewalle, “Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications”., Kluwer Academic Publishers, 2000.

[9] B. I. Sighencea, R. I. Stanciu and C. D. Căleanu, „A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction”. Sensors, 21(22):7543, 2021.

[10] J. Ziegler, et al., „Making Bertha Drive—An Autonomous Journey on a Historic Route”. IEEE Intell. Transp. Syst. Mag., 6, 8–20, 2014.

[11] C. Guo, et al., „Cooperation between driver and automated driving system: Implementation and evaluation”. Transp. Res. Part F Traffic Psychol. Behav., 61, 314–325, 2019.

[12] F. M. Ortiz, et al., „Vehicle Telematics via Exteroceptive Sensors: A Survey”. arXiv:2008.12632, 2020.

[13] Yole Developpement. MEMS and Sensors for Automotive: Market & Technology Report. 2017. Disponibil online: https://bit.ly/2X5pL70 (accessed on 23 July 2021).

[14] J. Kocic, et al., „Sensors and sensor fusion in autonomous vehicles”, 2018.

[15] B. I. Sighencea, I. Rares Stanciu and C. Sorandaru, "Using the α-β-γ Filter to solve the Threshold Problem," IEEE EUROCON 2021 - 19th International Conference on Smart Technologies, pp. 45-50, 2021.

[16] B. I. Sighencea, R. I. Stanciu, C. Șorândaru and C. D. Căleanu, „The Alpha-Beta Family of Filters to Solve the Threshold Problem: A Comparison”. Mathematics, 10, 880, 2022.

[17] B. I. Sighencea, I. R. Stanciu and C. D. Căleanu, „D-STGCN: Dynamic Pedestrian Trajectory Prediction Using Spatio-Temporal Graph Convolutional Networks”. Electronics, 12, 611, 2023.

[18] B. I. Sighencea, R. I. Stanciu and C. D. Căleanu, "Pedestrian Trajectory Prediction in Graph Representation Using Convolutional Neural Networks," 2022 IEEE 16th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 2022, pp. 000243-000248.

[19] J. Amirian, et al., „Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories with GANs”. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2964–2972, 16–17 June 2019.

[20] A. Mohamed, et al., „Social-STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction”. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14412–14420, 16–18 June 2020.
[21] B. I. Sighencea, "Pedestrian Trajectory Prediction Based on Tree Method using Graph Neural Networks," 24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 245-249, 2022.
[22] S. Pellegrini, „You'll never walk alone: Modeling social behavior for multi-target tracking”. In Proceedings of the IEEE 12th International Conference on Computer Vision, pp. 261–268, 27 September–4 October 2009.
[23] A. Lerner, „Crowds by example”. Comput. Graph. Forum, 26, 655–664, 2007.
[24] A. Robicquet, et al. „Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes”. In Computer Vision–ECCV, Springer, Volume 9912, 2016.
[25] B. I. Sighencea, R. I. Stanciu and C. D. Căleanu, „A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction”. Sensors, 21(22):7543, 2021.