



S-CGRU: an Efficient Model for Pedestrian Trajectory Prediction

Zhenwei Xu, Qing Yu, Wushouer Slamu, Yaoyong Zhou and
Zhida Liu

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

August 29, 2023

S-CGRU: An Efficient Model for Pedestrian Trajectory Prediction^{*}

Zhenwei Xu^{1,2}, YuQing^{1,3}, Wushouer Slamu^{1,3}, Yaoyong Zhou^{1,2}, and Zhida Liu^{1,3}

¹ Xinjiang University, Urumqi , 830000, Xinjiang Uygur Autonomous Region, People's Republic of China

² School of Software, Xinjiang University, Urumqi, People's Republic of China
{zhenweixu, 107552104337}@stu.xju.edu.cn

³ School of Information Science and Engineering, Xinjiang University, Urumqi, People's Republic of China
473461624@qq.com, wushour@126.com, 107552103689@stu.xju.edu.cn

Abstract. In the development of autonomous driving systems, pedestrian trajectory prediction plays a crucial role. Existing models still face some challenges in capturing the accuracy of complex pedestrian actions in different environments and in handling large-scale data and real-time prediction efficiency. To address this, we have designed a novel Complex Gated Recurrent Unit (CGRU) model, cleverly combining the spatial expressiveness of complex numbers with the efficiency of Gated Recurrent Unit networks to establish a lightweight model. Moreover, we have incorporated a social force model to further develop a Social Complex Gated Recurrent Unit (S-CGRU) model specifically for predicting pedestrian trajectories. To improve computational efficiency, we conducted an in-depth study of the pedestrian's attention field of view in different environments to optimize the amount of information processed and increase training efficiency. Experimental verification on six public datasets confirms that S-CGRU model significantly outperforms other baseline models not only in prediction accuracy but also in computational efficiency, validating the practical value of our model in pedestrian trajectory prediction.

Keywords: Pedestrian Trajectory Prediction · Gated Recurrent Unit · Complex number Neural Network · Autonomous Driving.

1 Introduction

Trajectory prediction plays a vital role in many critical applications in today's society, such as autonomous driving[2, 5], robotic path planning, behavior recognition, security surveillance, and logistics management. Parsing human motion trajectories is a research focus that involves multiple disciplines such as mathematics, physics, computer science, and social science [4]. Current prediction

^{*} Supported by organization x.

methods are typically divided into experience/rule-based methods and deep learning-based methods. Experience or rule-based prediction methods have high interpretability[5–7], but they do not perform well in terms of data fitting ability, often resulting in limited prediction accuracy[11–13]. Deep learning-based methods can provide more accurate prediction results and can automatically adapt to changes in data, demonstrating strong robustness and generalization capabilities[1, 8–10]. However, deep learning-based methods also have their limitations. First, such methods typically require a large amount of data to achieve optimal performance. Secondly, the prediction results often lack interpretability. In addition, these methods have high computational resource requirements [14].

In this paper, we propose a brand-new model, the Complex Gated Recurrent Unit (CGRU) model, which aims to combine the advantages of complex numbers and Gated Recurrent Units (GRU) to achieve efficient and accurate trajectory prediction. We leverage the powerful expressive ability of complex neural networks to capture complex patterns in complex space[16–20], thereby enabling the CGRU model to handle intricate trajectory prediction problems while maintaining high computational efficiency. On the other hand, by adopting the efficient simplicity of Gated Recurrent Unit networks [24], our model can effectively handle time-series data and model long-term dependencies. This model also considers the field of view of pedestrians walking in different environments, thus reducing the amount of information that the model needs to process and improving its operational efficiency. In summary, the main contributions of this paper can be encapsulated as follows:

- Innovatively proposed a Complex Gated Recurrent Unit (CGRU) model based on the concept of complex numbers. This deep neural network model combines the advantages of complex numbers in spatial representation and the efficiency of Gated Recurrent Unit networks, thereby constructing a lightweight and efficient Complex Gated Recurrent Unit model (CGRU).
- By integrating the social force model, we designed an S-CGRU model specifically for predicting pedestrian trajectories. This represents a significant extension and enhancement of traditional pedestrian trajectory prediction models.
- We conducted an in-depth study of the field of view of pedestrians walking in different environments, which led to optimization in data processing. This has significantly improved the training efficiency of our model.
- Empirical research results have verified the superiority of our model. Compared to other benchmark models, our model has shown excellent performance on six publicly available datasets, demonstrating its effectiveness and efficiency in pedestrian trajectory prediction.

2 Related Work

2.1 Complex-valued Neural Network

The use of complex numbers has shown many advantages in various fields, including computation, biology, and signal processing. From a computational

perspective, recent research on recurrent neural networks and early theoretical analyses have revealed that complex numbers can not only enhance the expressive power of neural networks but also optimize the memory extraction mechanism for noise suppression [15]. An increasing number of studies are focusing on complex number-based representations because they can improve optimization efficiency [16], enhance generalization performance [17], accelerate learning rates [18–20], and allow for noise-suppression memory mechanisms [19]. Studies in [20] and [18] show that the use of complex numbers can significantly improve the representational capabilities of Recurrent Neural Networks (RNNs).

From a biological perspective, Reichert and Serre [21] constructed a theoretically sound deep neural network model that achieves richer and more universal representations through the use of complex neuron units. This marks the first application of complex representations in deep networks. Conventionally, if the network data is entirely real numbers, we can only describe the specific numerical value of the intermediate output. However, once the data is entirely complex, in addition to being able to express the numerical magnitude of the intermediate output (the modulus of the complex number), we can also introduce the concept of time (the phase of the complex number). In neural network models, especially models like Gated Recurrent Units (GRU), the introduction of complex representations can greatly optimize the ability to process periodic changes and frequency information, as these pieces of information can be encoded through the phase of the complex number. Neurons with similar phases work in synchrony, allowing for constructive superposition, while neurons with different phases will undergo destructive superposition, causing interference with each other. This helps to differentiate and effectively manage the flow of information at different time steps, thereby enhancing the network’s efficiency and accuracy in handling time series data [40].

2.2 Recurrent Neural Networks

Recurrent Neural Networks (RNN) and its derivatives, such as Long Short-Term Memory (LSTM) networks [1] and Gated Recurrent Units (GRU) [24], are widely used for sequence prediction tasks. These RNN-based models have achieved remarkable results in fields like machine translation [25] and speech recognition [26]. RNNs are capable of capturing observed sequence patterns and generating predictive sequences based on these patterns.

The Gated Recurrent Unit (GRU) has unique advantages in dealing with these problems. Compared to LSTM, the structure of GRU is simpler, with fewer parameters and more efficient computation. In addition, the gating mechanism of GRU allows it to excel in capturing long-distance dependencies [25]. Therefore, GRU is widely applied in various sequence prediction tasks, including pedestrian trajectory prediction. However, current GRU models still require further research and improvements to better understand and handle interactions between pedestrians.

2.3 Human-human interactions

Since the beginning of the research, researchers have recognized that the influence of surrounding neighbors must be fully considered when predicting the future dynamic behavior of agents. As early as the initial stage, some work has first introduced the concept of social forces [30], describing their social interactions by simulating the repulsion and attraction between pedestrians. Subsequent research [31] introduced factors of personal attributes, calculating the impact of stationary crowds on moving pedestrians by classifying pedestrians. Furthermore, S-LSTM [1] successfully integrated the original LSTM network with the time step collection mechanism to simulate the social interaction of pedestrians. Since the social influence between pedestrians mainly depends on their distance from neighbors, research [32] proposed constructing a circular occupancy map to capture the influence of other pedestrians.

In our research, we adopted a data-driven method proposed by [1] to learn the interaction between people more deeply. Then, we expanded on this basis and further explored the field of view that pedestrians pay attention to during their movement. This method can more accurately simulate the behavior of pedestrians in real life because the direction and speed of pedestrian movement are often affected by objects and other pedestrians within their field of view. In addition, by reducing the irrelevant information that the model needs to process, it can improve the prediction accuracy and computational efficiency of the model, making it more practical in actual applications.

3 Method

3.1 Problem Definition

This paper aligns with the works of [1] and [8], assuming that there are N pedestrians during the prediction time period $[1, T_{pred}]$. After preprocessing the pedestrian trajectories in the video, the position of each pedestrian i at each time step t can be defined as a pair of spatial coordinates $P_t^i = (x_i^t, y_i^t)$, where $t \in \{1, 2, 3, \dots, t_{pred}\}$ and $i \in \{1, 2, 3, \dots, N\}$. Then, the coordinates of each pedestrian in the scene are divided into past trajectories X_i and future trajectories Y_i . As shown in Equation 1 and Equation 2:

$$X_i = \{P_t^i \mid t = 1, 2, 3, \dots, t_{obs}\} \quad (1)$$

$$Y_i = \{P_t^i \mid t = t_{obs} + 1, t_{obs} + 2, t_{obs} + 3, \dots, t_{pred}\} \quad (2)$$

Finally, this paper takes the past trajectories $\{X_i \mid i = 1, 2, 3, \dots, N\}$ of pedestrians in the scene as input, with the objective of generating future trajectories $\{\widehat{Y}_i \mid i = 1, 2, 3, \dots, N\}$ that closely resemble the actual future trajectories $\{Y_i \mid i = 1, 2, 3, \dots, N\}$ of the pedestrians. The generated future trajectories are defined as in Equation 3:

$$\widehat{Y}_i = \{\widehat{P}_t^i \mid t = t_{obs} + 1, t_{obs} + 2, t_{obs} + 3, \dots, t_{pred}\} \quad (3)$$

3.2 Architecture Overview

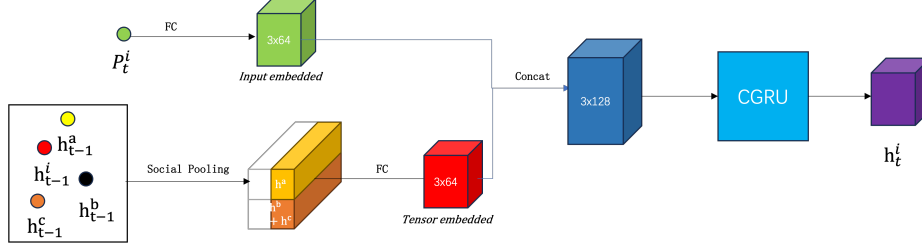


Fig. 1. At time step t , we integrate the neighbor information from the previous step and the current position of pedestrian i , update and obtain the current hidden state of pedestrian i .

Fig. 1 is a schematic diagram of the data flow of our model at time t . At time t , the inputs to the S-CGRU include the coordinates of pedestrian i , $P_t^i = (x_i^t, y_i^t)$, as well as the hidden state information at time $t-1$ (For example h_{t-1}^i). We use a social pooling layer to acquire the hidden state information of other pedestrians surrounding pedestrian i at time t (For example $h_{t-1}^a, h_{t-1}^b, h_{t-1}^c$). The social pooling layer of the hidden states implicitly infers the behavior of the nearby crowd, thereby adjusting its own path prediction. These nearby pedestrians are also influenced by their surrounding environment, and their behavior may change over time [8]. Next, we stack the hidden tensor at time t with the input data and feed them into the CGRU. After the CGRU processing, we obtain the hidden state at time t and the corresponding prediction results. Our model is built on the foundation of S-LSTM [1] but further optimizes the information processing and prediction mechanism. We particularly focus on how to more effectively extract information from the behavior of surrounding pedestrians, and how to integrate this information into our prediction model. By using the CGRU, we can more effectively utilize historical and environmental information, thereby improving the accuracy of pedestrian trajectory prediction.

3.3 Complex Gated Recurrent Unit

The Gated Recurrent Unit (GRU) is a variant of the Recurrent Neural Network (RNN) and was proposed by Kyunghyun Cho et. al [33]. The GRU solves the problem of gradient vanishing or gradient explosion that traditional RNNs may encounter when dealing with long sequences. The GRU introduces a gating mechanism to regulate the flow of information. Specifically, the GRU has two gates, the update gate and the reset gate:

- The update gate determines to what extent new input information is received. If the value of the update gate is close to 1, then the old memory is mainly retained; if it is close to 0, then the new input is mainly used.

- The reset gate determines how to use the previous hidden state when calculating the new candidate hidden state. If the value of the reset gate is close to 1, then most of the old memory is retained; if it is close to 0, then the old memory is ignored.

The reset gate determines how to use the previous hidden state when calculating the new candidate hidden state. If the value of the reset gate is close to 1, then most of the old memory is retained; if it is close to 0, then the old memory is ignored. Compared with the Long Short-Term Memory network (LSTM), the structure of the GRU is more concise because it only has two gates and does not have a separate cell state [33]. This makes the GRU computationally more efficient and easier to train, while still retaining good performance on many tasks, making it a popular choice in many neural network architectures.

In the Complex Gated Recurrent Unit (CGRU), this paper goes beyond the traditional real-number structure, extending it to the complex domain. In this way, each element not only contains a real part but also an imaginary part. This extension allows the model to perform calculations in the complex space, thus fully leveraging the unique characteristics of complex operations and demonstrating outstanding performance when processing temporal information. As shown in Fig. 2. In CGRU, we pass the input and hidden states separately into the

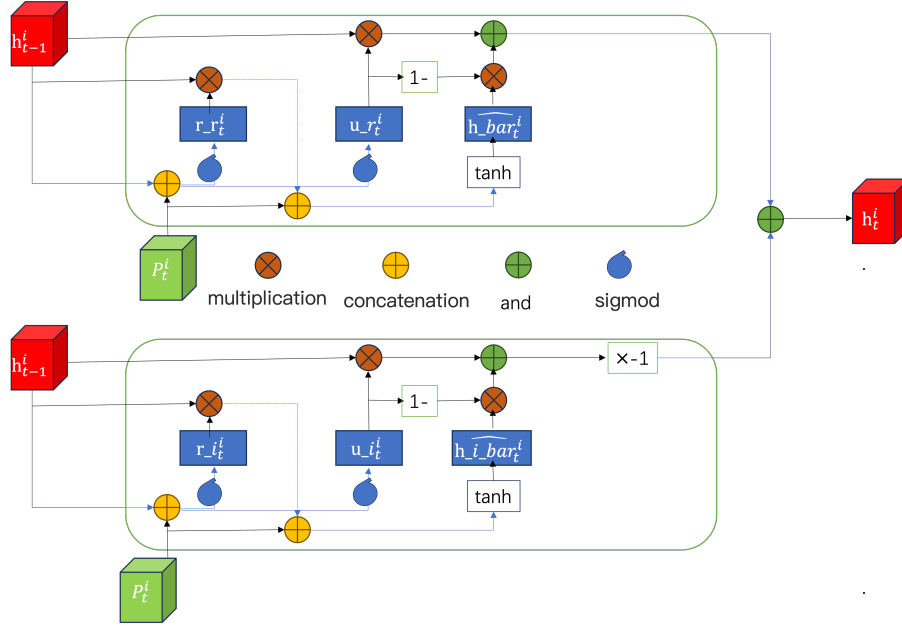


Fig. 2. Complex Gated Recurrent Unit.

linear transformation layers of the real and imaginary parts, and apply the sig-

moid activation function to obtain the real and imaginary parts of the reset gate and update gate. This enables calculations to be performed in the complex space. The characteristics of complex calculations can help the network remember important information in the long term, thereby improving the network's performance in dealing with long sequence problems.

In the reset gate, since the real and imaginary parts of the complex hidden state can encode information independently, the complex reset gate can selectively forget information in higher dimensions. This may allow the CGRU to perform better when dealing with complex, high-dimensional sequence data. The calculation formulas for the real and imaginary parts of the reset gate are shown in Equation 4 and Equation 5:

$$rr_t^i = \sigma(W_{ir} * P_t^i - W_{hr} * h_{t-1}^i) \quad (4)$$

$$ri_t^i = \sigma(W_{ii} * P_t^i - W_{hi} * h_{t-1}^i) \quad (5)$$

W_{ir} , W_{hr} , W_{ii} , and W_{hi} are the weights of the model, P_t^i is the input, and h_{t-1}^i is the previous hidden state. rr_t^i and ri_t^i are the real and imaginary parts of the reset gate, respectively.

The update gate in CGRU functions the same as in GRU, only now it operates on the complex form of the hidden state and the complex form of the input. This could possibly make CGRU perform better when handling complex, high-dimensional sequence data. The formulas for the update gate are shown in Equation 6 and Equation 7:

$$ur_t^i = \sigma(W_{uzr} * P_t^i - W_{hr} * h_{t-1}^i) \quad (6)$$

$$ui_t^i = \sigma(W_{uzi} * P_t^i - W_{hi} * h_{t-1}^i) \quad (7)$$

W_{uzr} , W_{hr} , W_{uzi} , and W_{hi} are the weights of the model, P_t^i is the input, and h_{t-1}^i is the previous hidden state. ur_t^i and ui_t^i are the real and imaginary parts of the update gate, respectively.

Based on the reset gate, we can calculate the real part of the new candidate hidden state $hbar_t^i$ (Equation 8) and the imaginary part $hibar_t^i$ (Equation 9). Equation 10 describes how to combine the update gate, the old hidden state, and the new candidate hidden state in the complex space to obtain the new hidden state h_t^i .

$$hbar_t^i = \sigma(W_{ir} * P_t^i - r_{rt}^i * W_{hr} * h_{t-1}^i) \quad (8)$$

$$hibar_t^i = \sigma(W_{ir} * P_t^i - r_{rt}^i * W_{hr} * h_{t-1}^i) \quad (9)$$

$$h_t^i = ur_t^i * h_{t-1}^i - ui_t^i * h_{t-1}^i + (1 - ur_t^i) * hbar_t^i - (1 - ui_t^i) * hibar_t^i \quad (10)$$

3.4 Exploring Pedestrians' Field of Vision in Different Environments

Understanding and considering the impact of other pedestrians within the field of vision is vital in predicting pedestrian motion patterns. As per studies, pedestrians typically adjust their trajectories based on the positions and movements

of other pedestrians within their field of view to avoid collisions. For other pedestrians behind an individual in certain environments, their influence on the trajectory is usually considered negligible. In this paper, we focus on exploring to what extent the movement trajectory of a pedestrian is affected by other pedestrians within a certain angular range. Our aim is to clarify and quantify this influence to accurately simulate this interpersonal interaction behavior when building predictive models. This approach not only helps to simplify the volume of information the model needs to process but also prevents unnecessary or irrelevant information from adversely impacting the accuracy of the prediction model. We hope that this method will enhance the accuracy of pedestrian motion trajectory predictions and provide a useful theoretical foundation and practical reference for related fields.

As shown in Figure 3, Figure 3a illustrates the scope of social pooling in [1, 8]: based on the hidden state of the LSTM encoder, it uses the method of maximum pooling; then, it merges the relative position coordinates of each neighbor to simulate the interaction between individuals. In models [1, 8], taking Figure 3a as an example, the emphasis is on analyzing the influence of other pedestrians on the trajectory of pedestrian P0 within a circle centered on the coordinates of P0 and with a certain distance as the radius. In Figure 2a, pedestrians P1, P2, P3, and P4 are all within this range, and they will affect the movement trajectory of P0. Conversely, since P5 is outside this range, we believe that P5 does not have a direct impact on the movement trajectory of P0.

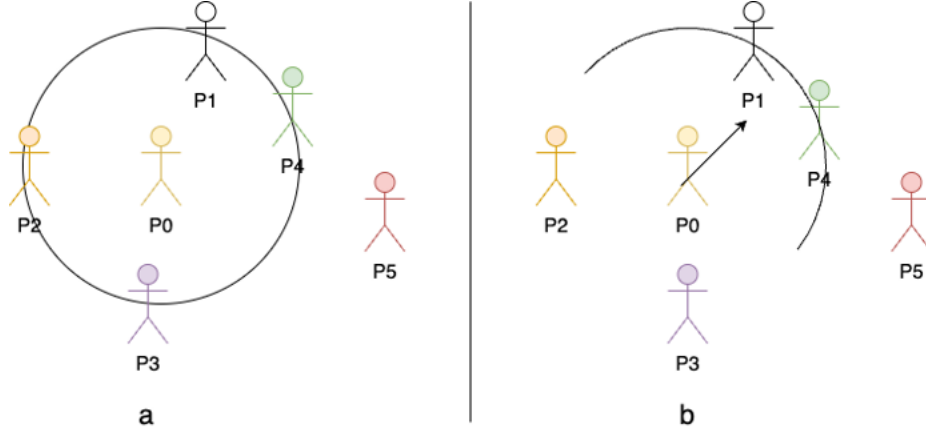


Fig. 3. Figure 3 illustrates the pedestrian movement pooling range. Figure 3a is a demonstration of the social pooling range based on [1, 8], while Figure 3b is a schematic diagram of the social pooling range proposed in this paper.

In the process of handling interactions, although pooling all the information of neighboring pedestrians seems to be a solution, we do not recommend this approach. The global information containing all pedestrians' data often contains

a large amount of redundant or irrelevant data. For instance, the movement information of pedestrians within a certain angular range behind the target pedestrian may have a negative impact on the accuracy of prediction results [4].

Therefore, we propose a new method that more comprehensively considers various factors that may affect pedestrian interactions. This method pays particular attention to the field of view (FoV) of the pedestrian, only considering those within this FoV as influences on the pedestrian’s future movements. Our assumption is based on the fact that people generally pay more attention to the events in their forward direction and ignore those behind. This effectively reduces the volume of unnecessary information and increases the computational efficiency of our model, making it more applicable to real-world scenarios.

In our approach, we set a hypothesis: other pedestrians who influence the trajectory of the target pedestrian should be located within a sector region forming a certain angle in front of the target pedestrian. We also strive to find the optimal angle of this sector region. Taking Figure 3b as an example, we explore how other pedestrians within a certain radius of the sector region (centered on pedestrian P0) influence P0’s movement trajectory. Pedestrians in this area, such as P1 and P4, will affect the movement path of P0. However, for pedestrians outside the region, such as P2, P3, and P5, we believe they will not have a direct impact on P0’s movement path. We further discussed the method to determine the optimal angle of the sector region in different scenarios. More detailed information can be found in Section 4.6.

3.5 Loss Functions

Similar to the method of the loss function in reference [34], the goal of training our model is to minimize the negative log-likelihood loss, as shown in Equation 11.

$$L = - \sum_{t=t_{obs}+1}^{t_{pred}} \log \{P(x_t^i, y_t^i | u_t^i, \sigma_t^i, \rho_t^i)\} \quad (11)$$

4 Experiments

4.1 Dataset and Evaluation Metrics

In the task of human trajectory prediction, this paper adopts six widely used datasets: Stanford Drone Dataset (SDD) [34], ETH datasets: Hotel and ETH [36], UCY datasets: UNIV, Zara1, and Zara2 [37]. The Stanford Drone Dataset (SDD) records videos within a university campus, containing six types of agents and rich interaction scenarios. In this dataset, there are approximately 185,000 interactions between agents and around 40,000 interactions between agents and the environment. The ETH/UCY datasets contain human trajectories from five scenarios, all of which record the movement paths of pedestrians in the world coordinate system. These datasets provide us with a variety of scenarios and rich

interactions, which are beneficial to our comprehensive study and understanding of human behavioral trajectories. The use of these datasets offers a broad sample for our research, helping us to understand and predict human behavioral trajectories more deeply.

In line with previous studies [1, 8], this paper uses two main metrics to evaluate model performance: 1) Average Displacement Error (ADE), 2) Final Displacement Error (FDE). ADE is a method to evaluate the mean square error between the predicted trajectory and the actual trajectory, which is calculated in the real-world coordinate system. It quantifies the average difference between the model’s predicted trajectory and the actual trajectory, thereby reflecting the overall prediction accuracy of the model (Equation 12.). On the other hand, FDE focuses on the endpoint of the predicted trajectory. It calculates the L2 distance between the last point of the predicted trajectory and the last point of the actual trajectory in the real-world coordinate system. FDE measures the model’s accuracy in predicting the endpoint of the trajectory. These two evaluation metrics can effectively evaluate the model’s overall performance in trajectory prediction and the accuracy of endpoint trajectory prediction (Equation 13).

$$ADE = \frac{\sum_{t=t_{obs}+1}^{t_{pred}} \|p_t^i - \hat{p}_t^i\|}{t_{pred} - t_{obs}} \quad (12)$$

$$FDE = \|p_{t_{pred}}^i - \hat{p}_{t_{pred}}^i\| \quad (13)$$

In order to maximize the use of these datasets and optimize our model, we adopt a special form of cross-validation known as the leave-one-out method [38]. Our approach is to train and validate the model on five datasets, then test on the remaining one. This process is iterated on each of the six datasets. Similarly, we also apply this training and testing strategy to the baseline methods used for model comparison.

4.2 Baselines

In order to evaluate the effectiveness of our proposed S-CGRU model, we compared it with some baseline methods for trajectory prediction, specifically as follows:

- Linear Model (Lin.): We use a pre-set Kalman filter, based on the assumption of linear acceleration to extrapolate trajectories.
- LSTM: LSTM is applied to cyclically predict each pedestrian’s future location from historical locations.
- S-LSTM [1]: This is a trajectory prediction model that combines LSTM and social pooling layer, capable of integrating the hidden states of neighboring pedestrians.
- S-GRU: This is a novel pedestrian trajectory prediction model based on GRU that we propose. This model serves as a baseline for validating the superior performance of the Complex Gate Recurrent Unit (CGRU) that we designed.

The comparison of the above four methods helps to comprehensively evaluate the performance of our proposed S-CGRU model.

4.3 Evaluations on ADE/FDE metrics

During the testing phase, we tracked trajectories for 3.2 seconds and predicted their paths for the next 4.8 seconds. At a frame rate of 0.4 seconds, this is equivalent to observing 8 frames and predicting the next 12 frames, which is similar to the setting in reference [8]. In 1, the performance of our model is compared with the baseline methods.

Table 1. Results on ETH , UCY and SDD based standard-sampling. 20 samples are used in prediction and the minimal error is reported.

Methods	Lin		LSTM		S-LSTM [1]		S-GRU (ours)		S-CGRU (ours)	
Metrics	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
ETH	1.31	1.35	0.96	1.36	0.82	1.11	0.96	1.24	0.54	0.73
Hotel	1.98	2.20	0.76	1.36	0.56	0.95	0.78	1.10	0.49	0.61
UNIV	0.71	0.62	0.65	0.68	0.34	0.42	0.56	0.83	0.34	0.39
ZARA1	1.20	1.76	1.09	1.84	0.56	0.95	0.76	1.12	0.49	0.59
ZARA2	0.72	0.98	0.82	1.18	0.40	0.54	0.65	0.73	0.39	0.44
SDD	0.83	1.12	0.62	0.93	0.50	0.75	0.60	0.82	0.48	0.63
AVG	1.13	1.34	0.82	1.23	0.53	0.79	0.72	0.97	0.46	0.57

From the results in 1, we can clearly see that our proposed S-CGRU model significantly outperforms the baseline model S-LSTM on all metrics. This fully validates the effectiveness of our proposed model in pedestrian trajectory prediction tasks. Firstly, compared to S-LSTM, our S-CGRU model has significantly improved in terms of prediction accuracy. This suggests that by utilizing Complex Gate Recurrent Units (CGRU), S-CGRU is able to more accurately capture and predict pedestrian movement trajectories. This can be attributed to CGRU’s excellent performance in handling sequential data, especially in capturing long-term dependencies. Secondly, the S-CGRU model offers better generalization performance. Regardless of the scenario or the number of pedestrians, S-CGRU can consistently output high-quality predictions. This indicates that our model can adapt well to various different circumstances, showing good adaptability to various environments. Finally, our S-CGRU model has excellent real-time performance. Compared to S-LSTM, even though both are close in computational complexity, our optimizations to the model structure allow S-CGRU to generate prediction results faster in actual applications, thereby meeting the needs of real-time applications. In summary, our S-CGRU model outperforms S-LSTM in pedestrian trajectory prediction tasks in terms of prediction accuracy, generalization ability, and real-time performance, overcoming the limitations of traditional GRU models. This validates the superiority of our model design and its application potential in pedestrian trajectory prediction tasks.

4.4 Model parameter amount and inference speed

To accurately assess the inference speed of various models, we specifically conducted an in-depth comparison of our Social-CGRU model and other publicly available models that can serve as comparison benchmarks. This comparison considers two key indicators of the model: parameter scale and inference speed. Test data is derived from time series data, obtained by densely sampling with a time step set to 1, and window size to 20 (including observation period Tobs(8) and prediction period Tpred(12)). When testing inference speed, we calculate the average inference time for all data segments.

Through such testing and comparison, we found that the S-CGRU model showed significant advantages. Since the S-CGRU model only contains two gates, reset and update, the number of its parameters is drastically reduced. Fewer parameters mean fewer variables need to be adjusted during the training process, which can significantly speed up model training. Moreover, fewer parameters make the S-CGRU model more lightweight, making it more suitable for running in resource-constrained environments, thereby improving its practicality. More importantly, the S-CGRU model has excellent performance in terms of inference speed. Fast inference speed not only provides faster feedback in real-time applications but also shows that the model can maintain high efficiency when handling large amounts of data, which helps improve overall work efficiency. This is especially important in application scenarios where quick decisions are needed or large amounts of data need to be processed. In summary, the excellent performance of the S-CGRU model in terms of parameter quantity and inference speed gives it a wide range of potential and practical value in various application scenarios. Detailed results are shown in 2.

Table 2. Comparisons of parameter amount and inference speed on ETH , UCY and SDD datasets.

Methods	Parameters (k)	Speed (time/batch)
S-LSTM	264	2.2
S-CGRU(ours)	231	1.6

4.5 The range of visual focus of pedestrians while walking in different scenarios.

We conducted in-depth research on six datasets: ETH, Hotel, UNIV, ZARA1, ZARA2, and SDD, aiming to reveal the range of vision during pedestrian walking in different scenarios. These datasets cover a variety of environments, such as schools, hotels, streets, intersections, etc., providing us with a wealth of pedestrian behaviors and movement patterns. In our research, we systematically adjusted the angle of the pedestrian’s field of view to simulate the degree to which they might be influenced by other pedestrians around them. Then, through the

training evaluation system, we detected and recorded the performance of pedestrian trajectory prediction. This method allows us to infer the main field of view that pedestrians pay attention to during walking based on the changes in prediction performance. Detailed results are shown in 2.

Table 3. Optimal field of view angles for different scenes on ETH , UCY and SDD datasets.

Scene (Dataset)	Optimal Field of View Angle (Degrees)
Street (ETH)	85
University (UNIV)	135
Hotel (Hotel)	75
Intersection (SDD)	120

We found that in various environments, the range of pedestrian vision is influenced by the surrounding environment. For example, in crowded environments such as hotels and street scenes, the range of pedestrian vision may narrow, mainly focusing on obstacles or other pedestrians directly in front. In contrast, in open environments like intersections, the range of pedestrian vision may expand as they need to consider information from more directions. Additionally, we discovered that the direction of pedestrian movement, speed, and individual characteristics (such as vision conditions) may also affect their field of view.

These findings have significant implications for our model design. Firstly, understanding the range of pedestrian vision can help us better understand and predict their behavior. Secondly, based on understanding the field of view, we can design more accurate and practical pedestrian trajectory prediction models. By considering the range of pedestrian vision, our model can make more precise predictions about pedestrian behavior in different environments, thereby improving the accuracy of prediction results.

5 Conclusions and future works

This research primarily explores the application of Complex Gated Recurrent Units (CGRU) in pedestrian trajectory prediction and proposes a new S-CGRU model, effectively integrating interpersonal interaction information and scene information. We first detailed the basic concepts and theories of complex neural networks and Gated Recurrent Units (GRU) and explored the advantages of using complex parameters in neural network models. Subsequently, we investigated the range of vision that pedestrians pay attention to during walking, providing a new theoretical perspective for understanding pedestrian behavior. Our research results show that the S-CGRU model has a significant advantage over the baseline model S-LSTM in dealing with pedestrian trajectory prediction issues. In the future, we will focus on person-scene interaction modeling (the interaction between pedestrians and their surroundings, and between pedestrians and vehicles). We hope to improve prediction performance by combining person-scene

interaction modeling with our S-CGRU model. Moreover, predicting pedestrian trajectories by combining pedestrian movement intentions is also a challenge that we need to address.

6 Acknowledgements

This work is supported by the National Natural Science Foundation of China (61562082), the Joint Funds of the National Natural Science Foundation of China (U1603262), and the "Intelligent Information R&D Cross-disciplinary Project" (Project Number: 202104140010). We thank all anonymous commenters for their constructive comments.

References

1. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S.: Social lstm: Human trajectory prediction in crowded spaces. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 961–971 (2016)
2. Deo, N., Trivedi, M.M.: Convolutional social pooling for vehicle trajectory prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 1468–1476 (2018)
3. Rudenko, A., Palmieri, L., Herman, M., Kitani, K.M., Gavrila, D.M., Arras, K.O.: Human motion trajectory prediction: A survey. *The International Journal of Robotics Research* 39(8), 895–935 (2020)
4. Jiangbei Yue, Dinesh Manocha, and He Wang. Human trajectory prediction via neural social physics. *arXiv preprint arXiv:2207.10435*, (2022).
5. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. *Physical review E* 51(5), 4282 (1995)
6. van den Berg, J., Lin, M., Manocha, D.: Reciprocal velocity obstacles for real-time multi-agent navigation. In: 2008 IEEE International Conference on Robotics and Automation (2008)
7. He, F., Xia, Y., Zhao, X., Wang, H.: Informative scene decomposition for crowd analysis, comparison and simulation guidance. *ACM Transaction on Graphics (TOG)* 4(39) (2020) 51(5), 4282 (1995)
8. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., Alahi, A.: Social gan: Socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2255–2264 (2018)
9. Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezaeifighi, H., Savarese, S.: Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1349–1358 (2019)
10. Mangalam, K., An, Y., Girase, H., Malik, J.: From goals, waypoints & paths to long term human trajectory forecasting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15233–15242 (2021)
11. Van Toll, W., Pettr'e, J.: Algorithms for Microscopic Crowd Simulation: Advancements in the 2010s. *Computer Graphics Forum* 40(2) (2021)

12. Wolinski, D., J. Guy, S., Olivier, A.H., Lin, M., Manocha, D., Pettr'e, J.: Parameter estimation and comparative evaluation of crowd simulations. *Computer Graphics Forum* 33(2), 303–312 (2014)
13. He, F., Xia, Y., Zhao, X., Wang, H.: Informative scene decomposition for crowd analysis, comparison and simulation guidance. *ACM Transaction on Graphics (TOG)* 4(39) (2020)
14. R. Korbmacher and A. Tordeux, "Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-Based Approaches," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24126–24144, Dec. 2022
15. Y. Bengio, and C.J. Pal. Deep complex networks. In *ICLR*, (2018)
16. T Nitta. On the critical points of the complex-valued neural network. In *Neural Information Processing*, (2002).
17. Akira Hirose and Shotaro Yoshida. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Transactions on Neural Networks and learning systems*, 23(4): 541–551, (2012).
18. Martin Arjovsky, Amar Shah, and Yoshua Bengio. Unitary evolution recurrent neural networks. *arXiv preprint arXiv:1511.06464*, (2015)
19. Ivo Danihelka, Greg Wayne, Benigno Uria, Nal Kalchbrenner, and Alex Graves. Associative long short-term memory. *arXiv preprint arXiv:1602.03032*, (2016).
20. Scott Wisdom, Thomas Powers, John Hershey, Jonathan Le Roux, and Les Atlas. Full-capacity unitary recurrent neural networks. In *Advances in Neural Information Processing Systems*, pages 4880–4888, (2016).
21. David P Reichert and Thomas Serre. Neuronal synchrony in complex-valued deep networks. *arXiv preprint arXiv:1312.6115*, (2013).
22. Rupesh K Srivastava, Klaus Greff, and Jürgen Schmidhuber. Training very deep networks. In *Advances in neural information processing systems*, pages 2377–2385, (2015).
23. Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, (2014).
24. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* 9(8), 1735–1780 (1997)
25. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *3rd International Conference on Learning Representations*. (2015)
26. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations*. (2017)
27. Antonini, G., et al.: Discrete choice models of pedestrian walking behavior. *Transport. Res. B* 40(8), 667–687 (2006)
28. Bahdanau, D., et al.: Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations*. (2015)
29. Lerner, A., et al.: Crowds by example. *Comput. Graphics Forum.* 26, 655–664 (2007)
30. D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 51, no. 5, p. 4282, 1995.
31. S. Yi, H. Li, and X. Wang, "Understanding pedestrian behaviors from stationary crowd groups," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3488–3496.

32. H. Xue, D. Q. Huynh, and M. Reynolds, "SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction," in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2018, pp. 1186-1194.
33. K. Cho, B. van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In EMNLP, (2014).
34. A. Mohamed, K. Qian, M. Elhoseiny, C. Claudel, Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction, in: CVPR, (2020).
35. Robicquet, A., Sadeghian, A., Alahi, A., Savarese, S.: Learning social etiquette: Human trajectory understanding in crowded scenes. In: European conference on computer vision. pp. 549–565. Springer (2016)
36. Pellegrini, S., Ess, A., Gool, L.V.: Improving data association by joint modeling of pedestrian trajectories and groupings. In: European conference on computer vision. pp. 452–465. Springer (2010)
37. Lerner, A., Chrysanthou, Y., Lischinski, D.: Crowds by example. In: Computer graphics forum. vol. 26, pp. 655–664. Wiley Online Library (2007)
38. Haowen Tang, Ping Wei, Jiapeng Li, and Nanning Zheng, "Evostgat: Evolving spatio-temporal graph attention networks for pedestrian trajectory prediction," Neurocomputing, vol. 491, pp. 333–342, (2022).
39. Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezaatofghi, H., Savarese, S.: Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1349–1358 (2019)
40. I. Danihelka, G. Wayne, B. Uria, N. Kalchbrenner, and A. Graves. Associative long short-term memory. In ICML, (2016).