Proceedings of the IDETC-CIE 2024
International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC-CIE 2024
August 25–28, 2024, Washington, DC, USA

IDETC-CIE2024

SAFER AND EFFICIENT FACTORY BY PREDICTING WORKER TRAJECTORIES USING SPATIO-TEMPORAL GRAPH ATTENTION NETWORKS

Satya Saravan Kumar K [1], Gokula Vasantha [1], Jonathan Corney [2], Jack Hanson [2], John Quigley [3], Hanane El-Raoui [3], Nathan Thompson [3], Andrew Sherlock [4]

[1] School of Computing, Engineering and the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, UK
[2] School of Engineering, The University of Edinburgh, Edinburgh EH8 9YL, UK
[3] Department of Management Science, University of Strathclyde, Glasgow G1 1XQ, UK
[4] National Manufacturing Institute Scotland, University of Strathclyde, PA49LJ, UK

ABSTRACT

Occupational accidents in manufacturing industries pose a significant risk, necessitating advanced strategies to ensure worker safety and enhance operational productivity. The unpredictable nature of worker movements, influenced by varied tasks such as material transportation, machine operation, and collaborative efforts, highlights the critical need for effective trajectory prediction mechanisms. This paper introduces an innovative approach utilizing Spatio-Temporal Graph Attention Networks (STGAT) and Spatio-Temporal Graph Convolutional Neural Networks (STGCNN) to predict worker trajectories with high accuracy and to analyze worker interactions within the manufacturing environment. Our methodology employs qualitative evaluation techniques to reveal intricate worker dynamics during assembly line processes, offering new perspectives on spatial-temporal interplays in a factory setting. By applying this method to movement data from a detailed case study involving six workers on a tribike assembly line, we demonstrate the effectiveness of our proposed algorithm in real-world scenarios. The utilization of advanced Graph Neural Network technologies allows for the precise modeling of complex spatial-temporal relationships, enabling the accurate prediction of worker paths. This research contributes significantly to the fields of occupational safety and industrial efficiency by providing a comprehensive framework for anticipating worker movements and understanding their interactions in intricate manufacturing landscapes. Moreover, it addresses existing challenges in trajectory prediction and outlines potential directions for future research, aiming to broaden the application of predictive analytics in enhancing safety protocols and operational strategies in the manufacturing sector.

Keywords: Human motion trajectory prediction, Graph Neural Networks, Smart factory, Engineering Informatics, Intelligent Manufacturing.

1. INTRODUCTION

In the realm of manufacturing industries, the safety and efficiency of factory operations are paramount. The dynamic interplay of machinery, human workers, and logistical processes presents a complex environment where occupational accidents are a significant concern. Human workers, central to the production process, navigate these environments, making spontaneous decisions for material transfer, collaboration, and information exchange. This unpredictable aspect of human movement poses challenges to maintaining a safe and efficient workspace. Predicting worker movements within such an environment can offer substantial benefits, from enhancing safety protocols to optimizing the overall productivity of the manufacturing process. The urgency for such advancements is underscored by the need to prevent accidents and collisions, which can lead to significant downtime, financial loss, and, most importantly, injury to personnel.

This paper introduces a novel approach to worker trajectory prediction within manufacturing settings, focusing on improving safety measures and operational efficiency. Traditional methods often fall short in addressing the dynamic and interactive nature of factory environments. Our work leverages Spatio temporal Graph Attention Network models to predict worker movements and worker interactions using the attention weights. By incorporating these attention weights, we can conduct a qualitative evaluation of various working scenarios, such as instances when all workers are dwelling, a mix of workers in transition and dwelling, and situations where workers are nearby or further apart. This evaluation facilitates the classification and creation of an event log of workers based on position data, which is pivotal for incident reporting, safety monitoring, process optimization and understanding the flow of human activity in complex industrial spaces. The research contributes to establishing a baseline trajectory prediction performance measure for a product assembly factory using state-of-the-art prediction approaches.

Illustrated through an empirical study of six workers in a tribike assembly line, our findings demonstrate the potential of our proposed method in real-world applications. The analysis not only showcases the accuracy of our predictions but also sheds light on the intricacies of human movement within a manufacturing context. Through this exploration, we aim to contribute to the ongoing efforts to create safer and more productive factory environments, addressing both the immediate challenges of occupational safety and the broader objectives of industrial efficiency. Furthermore, by modeling and predicting worker paths, our approach facilitates the design of more efficient workflows, reducing unnecessary movement and optimizing the layout of physical spaces for smoother operations.

In the following sections, we delve into the methodology behind our prediction algorithm, discuss the data collection process in the assembly line scenario, present our findings, and explore the implications of our research for future technological and process advancements in the field of manufacturing.

2. RELATED LITERATURE

2.1 Related literature on worker's movement prediction

In the advanced manufacturing industry, predicting workers' movement is paramount for operational efficiency, considering the integrated environment of products, tools, machines, processes, and human workers. Designing effective movement systems within the factory largely depends on human awareness. Compared to path planning, which aims to find the optimal paths between destinations, trajectory predictions foresee future trajectories by learning from past movement positions. This worker's trajectory prediction has been helpful for multiple applications, including adjustable route planning for automated guided vehicles, congestion elimination, collision detection, and process scheduling. The path dynamics between humans and robots are widely discussed in the literature. Multiple literature surveys exist on human motion trajectory prediction [1-7]. This section discusses the importance of trajectory prediction, and research focusing on manufacturing and other domains.

2.2 Importance of Trajectory Prediction

Understanding uncertainties of human behaviour through human movement trajectory has wider applications. Aliabadi et al. [8] identified that safety policies and procedures and their understanding play a vital role in predicting accidents and influencing a safer climate in an organization. Trajectory predictions could develop an effective safety procedure based on workers' movement observations and behavioural changes. Löcklin et al. [9] proposed a Bayes classifier to predict the probability of the next destination utilising manufacturing schedule and real-time position data. They argued that human motion intentions and abstract activity modelling (rather than a fixed point of interest) help increase prediction reliability up to 80% with a few training data sets due to restricting the number of probable destinations. However, prediction drops if the intentions are unclear. Although schedule awareness helps in path predictions, Ayse et al. [10] observed that workers significantly deviate from the assigned schedule based on task needs. So, the schedule awareness should also be dynamically updated to improve trajectory prediction.

Besides the manufacturing domain, human trajectory prediction is widely used in construction[11-12], warehouse [13], and emergency situations like fire evacuation [14]. Cai et al. [11] applied Deep Reinforcement Learning to predict the movements of construction workers that can be integrated to achieve safer path planning for robots. Although efficient path planning was achieved with this approach, the collision rate was reduced only by 23%. There is a significant scope for improvement in collision prevention. Similarly, Hu et al. [12] applied a context-aware Long Short-Term Memory (LSTM)-based method to predict worker's trajectory in construction robot path planning. Gong et al. [13] highlighted the importance of uncertainties of human behaviour in pickup and delivery task planning and proposed a human-swarm hybrid system to accomplish storage system tasks. Hong et al. [15] used a reinforcement learning method (deep Q-network) to find optimal path prediction to evacuate workers in case of an accident.

Also, providing a proactive warning for collision detection is vital to prevent accidents. Kim et al. [16] used deep neural networks to achieve an average proximity error of 0.95 m in predicting 5.28 s future proximity distance between a worker and a truck. Luo et al. [17] highlighted the importance of trajectory prediction in layout design and its influence on the efficiency of the goods-picking system. Li et al. [18] used a discrete-time Markov chain mathematical model to identify hazardous regions on the construction site using a real-time location system. Tang et al. [19] utilised a long short-term memory (LSTM) encoder-decoder in video frames to predict worker and equipment motion trajectories on construction sites. The results show average localisation errors of 7.30, 12.71, and 24.22 pixels for 10, 20, and 40 future steps, respectively. Task allocation based on trajectory prediction was also studied in mobile crowdsensing and crowdsourcing applications [20]. A longer prediction horizon is a major challenge in various studied domains.

2.3 Human Trajectory Prediction Algorithms

The conceptual framework for understanding human trajectory dynamics in groups was established by Helbing et al. [21] through the Social Force Model. This model introduced the idea of using attractive forces to guide individuals towards their goals while employing repulsive forces to maintain personal space and avoid collisions. Despite its foundational importance, the model's reliance on predetermined potential functions proved too simplistic to capture the complex interplay of social interactions. This limitation was also evident in subsequent models that extended the Social Force concept.

2.3.1 Deep Learning based Algorithms

The introduction of deep learning to the domain marked a paradigm shift. The pioneering Social-LSTM [1] utilized recurrent neural networks (RNNs) to model the motion of individuals within a crowd, aggregating these motions through a pooling mechanism to predict future trajectories. Subsequent enhancements, such as Peek into the Future (PIF) [22] and State-Refinement LSTM (SR-LSTM) [23], incorporated visual cues and innovative pooling strategies to refine prediction accuracy.

Deep learning combined with modelling social interactions has been instrumental in the evolution of human trajectory prediction. Earlier models like Social GAN [24] and Sophie [25] were some of the first to employ deep generative models and attention mechanisms, setting the stage for future developments. increasingly focused on accurately modeling social interactions within crowds. Techniques exemplified by Social-GAN [24] and Sophie emphasize the importance of considering the multi-modal nature of human movements. The Social-BiGAT [26] model employs graph attention networks to dynamically weigh the influence of individual trajectories, showcasing the shift towards more interactive and adaptive models.

2.3.2 Graph-Based Approaches

A significant advancement in the trajectory prediction field has been the adoption of graph-based models. This approach treats human trajectories and interactions as graphs, allowing for a more nuanced representation of social interactions. The development of Graph Convolutional Networks (GCNs) and their application across various domains paved the way for models like ST-GCNN [27] STGlow [28] which adapted graph convolutions for spatio-temporal data analysis.[29]

2.3.3 Transformer Models

Agent Former and Transformer TF [30] explored the use of transformer models, adapting the architecture for socio-temporal multi-agent forecasting, showcasing the versatility of transformers in capturing dynamic interactions.

PECNet [31] and Sophie [25] focused on endpoint conditioned prediction and attentive GAN frameworks, respectively, proposing novel pathways to address the prediction of socially and physically constrained paths.

2.3.4 State-of-the-art Models and Recent Breakthroughs

AMEND [32] and MemoNet [33] represent the cutting edge, incorporating mixture of expert's frameworks and memory-augmented networks to tackle the challenge of long-tailed trajectory predictions, indicating a shift towards more nuanced and context-aware modeling.

ExpertTraj [34] and Pishgu [35] introduce highly specialized approaches, with ExpertTraj offering dynamic prediction based on expert goal examples and Pishgu proposing a universal path prediction network for edge systems. Neural Social Physics (NSP) [36] presents a unique approach to human trajectory prediction by encapsulating social physics concepts within neural networks. It emphasizes understanding the underlying social forces that govern human movement in crowds.

Y-Net [37] takes inspiration from waypoints and paths, offering long-term human trajectory forecasting. It integrates goal-oriented behaviour with neural network capabilities to predict movements over extended periods accurately. V2-Net [38] employs a hierarchical structure for trajectory prediction, utilizing Fourier spectrums to capture both the spatial and temporal aspects of pedestrian movements. This method offers a novel perspective by looking at the problem vertically.

Although human motion movement is largely discussed in pedestrian's domain, the applications of latest motion prediction methods to the manufacturing domain are limited. As manufacturing environments continue to evolve towards greater automation and integration, the ability to accurately predict worker movements becomes increasingly crucial. The advancements highlighted in the literature point towards a future where trajectory prediction models not only enhance safety protocols and efficiency but also pave the way for innovative solutions to managing human-machine interactions. This evolution of prediction models holds the promise of creating more adaptive, safe, and efficient manufacturing ecosystems.

3. RESEARCH AIM AND METHODOLOGY

3.1 Problem Definition

There are multiple prediction problems, such as location-, position-, destination- and route prediction. With the taxonomy proposed by Rudenko et al. [1], compared to physics-based motion prediction, pattern- and planning-based approaches are widely used for prediction, considering the advancement of real-time Location Systems (RLTS). Also, the taxonomy emphasizes the importance of dynamic and static environment cues.

Based on the taxonomy our approach leverages route prediction modelling using sequential models, which are adept at capturing the time-continuity inherent in human movement. By implementing sequential models, we can use past trajectories to predict future locations, effectively modelling the movement dynamics of individuals and groups over time.

Furthermore, our methodology can also be incorporated to context-based dynamic environment cues, which are integral to understanding and anticipating human movement within a certain area. These cues encompass a range of factors, from the static environment, such as the layout of a factory floor, to dynamic elements like moving machinery or the presence of other workers. The interplay of these factors influences decision-making processes and subsequent movement patterns of workers, necessitating a model that can account for such complexities.

Group-aware agents form a crucial aspect of our model. Recognizing that workers do not operate in isolation, our system accounts for the collective behaviours and interactions within a group. By understanding social dynamics and the rules that govern group behaviours, the model can predict not only individual trajectories but also the collective movement of groups.

The research aims to develop a novel trajectory prediction approach that emphasizes integrating individual trajectories and the temporal dependencies between workers over time. We examine a dynamic scene with N workers, represented as $p_1, p_2, \ldots, p_N$. The position of a worker p_i (where i ranges from 1 to N) at a given time step t is denoted by $S_i^t = (x_i^t, y_i^t)$. Our dataset comprises the recorded positions S_i^t for each worker $i = 1,2, \ldots, N$ at discrete time steps $t = 1, \ldots, T_{\text{obs}}$. The goal is to predict the trajectory, i.e., the sequence of future positions S_i^t, for the forthcoming time steps $t = T_{\text{obs}} + 1, \ldots, T_{\text{pred}}$.

Let's denote the set of workers at time step t as $\{p_1^t, p_2^t, \ldots, p_N^t\}$, where each p_i^t is a vector representing the position of worker i in a graph structured data. The attention parameter α_{ij}^t represents the weight of the edge from worker j to worker i at time step t, which is calculated as:

$$\alpha_{ij}^t = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}h_i^t \;\|\; \mathbf{W}h_j^t\right]\right)\right)}{\sum_{k\in\mathcal{N}_i} \exp\left(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}h_i^t \;\|\; \mathbf{W}h_k^t])\right)} \qquad (1)$$

where:

- h_i^t is the feature vector of worker i at time step t,
- $\|$ denotes concatenation,
- $\mathbf{W}$ is a shared linear transformation applied to the feature vector of each worker,
- $\mathbf{a}$ is a weight vector of a single-layer feedforward neural network,
- $\mathcal{N}_i$ is the set of neighbors of worker i in the graph,
- LeakyReLU is an activation function with a small slope for negative values to allow a small gradient when the unit is not active,
- exp denotes the exponential function,
- α_{ij}^t is normalized across all choices of j using the softmax function to ensure that the weights sum to 1.

This attention mechanism allows the model to focus on certain parts of the input graph, which is crucial for learning complex dependencies between workers. By assigning these weights dynamically for each time step, the model can adapt the influence of the surrounding workers' positions on the predicted trajectory of a particular worker.

In this research, we explore the prediction of worker path movement within environments, such as factories or construction sites, leveraging advanced graph neural network architectures. Our methodology uses two models "STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction" [27] and "STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction" [39]. Specifically, we adapt the Graph Attention Network ($GATv_2$) [40] for enhanced performance and employ the second study's framework for qualitative analysis on our worker dataset. This section delineates the architecture, functional definitions, loss functions, graph implementations, and application of these models to our dataset.

3.2 Model Overview

3.2.1 STGAT Model for Worker Path Prediction

The model architecture from Huang et al. [39] forms the basis for our qualitative analysis, offering a robust mechanism to model spatial-temporal interactions among workers. The STGAT framework incorporates LSTM networks to encode individual worker trajectories [41] and GAT layers to capture the complex dynamics of worker interactions over time. This dual approach ensures a comprehensive analysis of both individual movement patterns and group dynamics, essential for accurate trajectory prediction. In our scenario, we consider N workers present within a factory scene, denoted as $\mathrm{n}_1, \mathrm{n}_2, \ldots, \mathrm{n}$. The position of worker $\mathrm{n_i} (\mathrm{i} \in [1, \mathrm{N}])$ at time-step t is represented as $\mathrm{P_i^t} = (\mathrm{x_i^t}, \mathrm{y_i^t})$.

To capture each worker's unique motion pattern, we employ an LSTM for each, termed M-LSTM (Motion Encoding LSTM). Initially, we calculate each worker's relative position change:

$$\Delta \mathrm{x_i^t} = \mathrm{x_i^t} - \mathrm{x_i^{t-1}}, \Delta \mathrm{y_i^t} = \mathrm{y_i^t} - \mathrm{y_i^{t-1}} \qquad (2)$$

These relative positions are then transformed into vectors $\mathrm{e_i^t}$ using an embedding function ϕ with weight $\mathrm{W_{ee}}$, and fed into the M-LSTM:

$$\mathrm{e_i^t} = \phi(\Delta \mathrm{x_i^t}, \Delta \mathrm{y_i^t}; \mathrm{W_{ee}}) \qquad (3)$$

$$\mathrm{m_i^t} = \mathrm{M - LSTM}(\mathrm{m_i^{t-1}}, \mathrm{e_i^t}; \mathrm{W_m}) \qquad (4)$$

Here $\mathrm{m_i^t}$ is the hidden state of the M-LSTM at time-step t, with $\mathrm{W_m}$ being the weight of the M-LSTM cell. These parameters are uniformly applied across all workers. The uniform application of parameters across all workers in the model serves several key purposes. Firstly, it ensures consistency in the representation and processing of motion patterns across different individuals within the same environment. This consistency is critical for the model to learn generalized patterns of motion that are not specific to any single worker but rather applicable to any worker within the work environment. By applying the same set of parameters, the model can effectively capture and learn from the collective dynamics of worker movements, which is essential for accurately predicting future trajectories in a diverse workforce.

The simplistic approach of using one LSTM per worker does not account for the complex interactions between workers in a crowded factory environment. To facilitate information sharing among workers and model human-human interactions effectively, we conceptualize the workers in a scene as nodes in a graph and employ a GAT mechanism for information aggregation. The GAT mechanism operates on graph-structured data, calculating the features of each graph node by attending over its neighbours through a self-attention strategy. By stacking graph attention layers, the GAT model can compute the output features $h' = \{h'_1, h'_2, \dots, h'_N\}$ from the input features $h = \{h_1, h_2, \dots, h_N\}$, where N is the number of nodes (workers) [39].

The hidden states m_i^t (for $t = 1, \dots, T_{obs}$) derived from each worker are introduced into the graph attention framework. Within this context, the interaction strength between any two workers (i, j) at a given time step t is quantified using attention coefficients, calculated using Eq. 1.

Upon normalization of the attention coefficients, the resultant output for node i at time t, following one graph attention layer, is articulated as:

$$\hat{m}_i^t = \sigma(\sum_{j \in N_i} \alpha_{ij}^t W m_j^t) \quad (5)$$

where σ represents an activation function. Equations pertaining to the computation of attention coefficients and subsequent output highlight the operational essence of a single graph attention layer. For our model, we incorporate two such layers, culminating in $\hat{m}_i^t$, which aggregates the hidden states post two layers of graph attention, encapsulating the spatial influences exerted by adjacent workers on worker i at time t.

To adequately capture worker interactions within densely populated factory environments, traditional LSTM-based frameworks often exchange hidden states between workers. Nonetheless, these approaches focus on simultaneous time-step interactions, omitting the intricate temporal dynamics that unfold between these interactions. Addressing this gap, we introduce a novel LSTM variant, dubbed G-LSTM, aimed at explicitly modelling the temporal dependencies that interlink worker interactions over time:

$$g_i^t = \text{G-LSTM}(g_i^{t-1}, \hat{m}_i^t; W_g) \quad (6)$$

Here, $\hat{m}_i^t$ is derived from Eq.5, representing the aggregated spatial information at time-step t, while W_g signifies the shared G-LSTM weight across all worker sequences.

Within the encoder phase, two distinct LSTM models (M-LSTM for individual motion patterns and G-LSTM for interaction-based temporal correlations) are employed. These models synergize to meld spatial and temporal data streams effectively. At the observation endpoint T_{obs} each worker is associated with dual hidden states $(m_i^{T_{obs}}, g_i^{T_{obs}})$ emanating from the respective LSTMs. Preceding concatenation, these states are individually processed through separate multilayer perceptron's (δ_1 and δ_2):

$$\bar{m}_i = \delta_1(m_i^{T_{obs}}) \quad (7)$$

$$\bar{g}_i = \delta_2(g_i^{T_{obs}}) \quad (8)$$

$$h_i = \bar{m}_i \parallel \bar{g}_i \quad (9)$$

Addressing the unpredictability of worker movements, we apply a variety loss to encourage diverse trajectory predictions. The model's intermediate state vector includes M-LSTM and G-LSTM hidden states, alongside added noise:

$$d_i^{T_{obs}} = h_i \parallel z \quad (10)$$

Predicted relative positions are derived using a decoder LSTM (D-LSTM):

$$d_i^{T_{obs}+1} = \text{D-LSTM}(d_i^{T_{obs}}, e_i^{T_{obs}}; W_d) \quad (11)$$

$$(\Delta x_i^{T_{obs}+1}, \Delta y_i^{T_{obs}+1}) = \delta_3(d_i^{T_{obs}+1}) \quad (12)$$

Here, W_d denotes the D-LSTM weight, and δ_3 is a linear transformation layer, with $e_i^{T_{obs}}$ stemming from Eq. 3. Subsequent D-LSTM inputs are iteratively adjusted based on the latest predicted relative positions, facilitating a seamless transition to absolute positioning for loss computation.

The variety loss, as delineated by [24], incentivizes the model to yield multiple trajectory forecasts per worker by random sampling of z from a standard normal distribution. Among these, the trajectory nearest to the actual path is selected for loss computation, promoting model versatility and adherence to historical movement patterns:

$$L_{variety} = \min_k \parallel Y_i - \hat{Y}_i^k \parallel^2 \quad (13)$$

This loss mechanism not only underscores the model's capacity to navigate the spectrum of possible outcomes but also aligns closely with our overarching goal of enhancing workplace safety and efficiency through precise and adaptable worker movement predictions.

3.2.2 STGCNN Model Architecture for Worker Path Prediction

We replace the original GAT with GATv2 in the Social-STGCNN framework to leverage its improved attention mechanism, which facilitates more effective learning of spatial dependencies among workers. The GATv2 version enhances the model's ability to focus on relevant nodes (workers) dynamically, improving the prediction accuracy of their future positions. This adaptation allows for a more nuanced understanding of social interactions and spatial-temporal correlations within crowded work environments. This model merges two primary components: the Spatio-Temporal Graph Convolution Neural Network (ST-GCNN) for feature extraction from worker trajectories, and the Time-Extrapolator Convolution Neural Network (TXP-CNN) for forecasting future movements.

We construct spatial graphs G_t for each time step t, symbolizing the positional interrelations of workers. Each graph

$G_t = (V_t, E_t)$ comprises vertices V_t representing workers and edges E_t indicating their interactions. The adjacency matrix A_t, weighted by kernel function outcomes $a_{ij,t}$ encapsulates the strength of these interactions.

ST-GCNN for Feature Extraction: Utilizing the graph representations, ST-GCNN performs spatio-temporal convolution operations to derive a compact set of features, encapsulating the historical trajectory information of all observed workers.

TXP-CNN for Trajectory Prediction: Building upon the extracted features, TXP-CNN extrapolates these spatio-temporal embeddings to predict collective future movements, leveraging convolutional operations for temporal expansion.

To facilitate learning, we normalize each temporal slice of the adjacency matrix A_t symmetrically, enhancing the efficacy of subsequent graph convolution operations. This normalization process employs the diagonal node degree matrix Λ_t to adjust A_t ensuring balanced feature propagation.

ST-GCNN layers aggregate and process the information from worker nodes and their interactions across both spatial and temporal dimensions. The resulting embeddings, denoted as $\bar{V}$, serve as a comprehensive representation of the workers' movement patterns and mutual influences over the observed time frame. Receiving the spatio-temporal node embeddings $\bar{V}$ from ST-GCNN, TXP-CNN manipulates the time dimension of these embeddings to predict future trajectories. It employs a series of convolutional layers, some with residual connections, to model the temporal progression of worker movements effectively. This component allows our model to address the variable nature of industrial workspaces, predicting multiple potential future paths for each worker.

The first layer of TXP-CNN, directly interfacing with ST-GCNN outputs, lacks a residual connection due to dimensional discrepancies between the observed and predicted samples. Subsequent layers incorporate residual connections to refine the prediction accuracy, ensuring a coherent transition from historical data to future trajectory forecasts.

3.3 Dataset and Performance Evaluation Metrics

Our study employs two well-regarded human trajectory prediction datasets: ETH and UCY, to train and evaluate the Social-STGCNN v2 and the STGATv2 models.

- ETH Dataset: This dataset comprises two distinct scenarios, ETH and HOTEL, offering diverse settings to test the models' robustness across different pedestrian behaviours and environments.
- UCY Dataset: Encompassing three scenarios named ZARA1, ZARA2, and UNIV, the UCY dataset provides additional variability with its inclusion of rich human-human interaction data and complex movement patterns.

Both datasets are sampled at intervals of 0.4 seconds, providing a detailed temporal resolution for trajectory analysis. In aligning with prior benchmark practices such as those employed in Social-LSTM, the models are trained on subsets of specific datasets and tested against the remaining data, validating across all five data subsets of both the ETH and UCY datasets for comprehensive evaluation.

Worker dataset: The dataset consists of 2D worker position data that has been collated through indoor localization sensors (UWB tags) used for tracking of worker movements during a three-hour work shift. [43]. The UWB data presents a consistent stream of position samples which are sampled at intervals of 1 second. Each worker is tagged, allowing for individual movement patterns to be analyzed. The data is time-stamped, enabling the study of movement over the course of the shift.

To assess the performance of our models, we employ two standard metrics widely recognized in trajectory prediction research: Average Displacement Error (ADE) and Final Displacement Error (FDE).[44]

Average Displacement Error (ADE): This metric calculates the mean Euclidean distance between the predicted positions and the actual ground truth positions across all points in the trajectory. It provides an overall measure of the model's accuracy throughout the prediction horizon.

$$ADE = \frac{\sum_{n \in N} \sum_{t \in T_p} \| \hat{p}_{nt} - p_{nt} \|_2}{N \times T_p} \tag{14}$$

Here, N represents the total number of pedestrians, T_p denotes the prediction time steps, $\hat{p}_{nt}$ is the predicted position, and p_{nt} is the ground truth position at time t or pedestrian n.

Final Displacement Error (FDE): This metric focuses on the Euclidean distance between the predicted and actual positions at the final time step T_p of the prediction horizon. It specifically evaluates the accuracy of the model's endpoint prediction.

$$FDE = \frac{\sum_{n \in N} \| \hat{p}_{nT_p} - p_{nT_p} \|_2}{N} \tag{15}$$

Both ADE and FDE are computed using the closest sample to the ground truth out of 20 samples generated from the predicted bi-variate Gaussian distribution. This approach aligns with the evaluation strategy adopted in seminal works like Social-LSTM and Social-GAN, facilitating a fair comparison across models and ensuring the evaluation reflects the models' capability to predict plausible pedestrian trajectories.

Table 1 ADE and FDE values for the models on respective datasets for T_{in} = 8 and T_{pred} =12 seconds.

Model/Datasets	ETH	HOTEL	UNIV	ZARA1	ZARA2	Worker
LSTM	1.09	0.86	0.61	0.41	0.52	1.16
	2.41	1.91	1.31	0.88	1.11	2.13
STGCNN	0.64	0.41	0.48	0.34	0.30	0.65
	1.11	0.68	*0.91*	0.53	0.48	0.70
STGAT	0.65	0.35	0.52	0.34	**0.29**	0.55
	1.12	*0.66*	1.10	0.69	0.60	0.76
STGCNN v2	0.72	0.41	**0.48**	**0.33**	0.30	0.64
	1.20	0.67	*0.91*	0.51	*0.48*	*0.68*
STGAT v2	**0.63**	**0.33**	0.51	0.35	0.31	**0.53**
	1.11	0.68	1.12	0.68	0.66	0.74

ADE	**FDE**

4. RESULTS AND DISCUSSION

4.1 Model Configuration and Training Setup

For the STGCNN v2 model, we opted for a training batch size of 128, conducting the training over 250 epochs. We employed Stochastic Gradient Descent (SGD) as the optimization technique to adjust the model weights. Initially, the learning rate was set to 0.01 to facilitate rapid convergence towards the global minimum of the loss function. Recognizing the need for fine-tuning as training progressed, we reduced the learning rate to 0.002 after the first 150 epochs. This strategy helped in achieving a balance between convergence speed and training stability, allowing the model to learn the nuances of the dataset effectively.

The training configuration for the STGAT v2 model was similarly tailored to the demands of our trajectory prediction task but with some variations to cater to the model's architecture and learning dynamics. We decided on a smaller batch size of 64 for STGAT v2, running the training for a total of 400 epochs to provide ample opportunity for the model to learn from the complex spatial-temporal relationships in the data. The Adam optimizer was chosen for its adaptive learning rate capabilities, starting with a learning rate of 0.01 to expedite the initial learning phase. To further refine the model's accuracy, we decreased the learning rate to 0.005 after 250 epochs, aiding in the fine-tuning of the model's parameters as it approached optimal performance. We utilized a data split ratio of 70% for training, 15% for validation, and 15% for testing.

4.2 Quantitative Analysis

The quantitative analysis of human path prediction models offers insights into their performance across various environments represented by the ETH, HOTEL, UNIV, ZARA1, ZARA2, and Worker datasets. We evaluated the models based on two key metrics: Average Displacement Error (ADE) and Final Displacement Error (FDE), which measure the prediction accuracy over entire predicted trajectories and at their final positions, respectively.

The depicted image Figure 1 illustrates the trajectory prediction for a worker within a manufacturing environment, showcasing three key components: the input path, the ground truth, and the model's prediction. The red line represents the worker's actual path leading up to a certain point in time, serving as the input for the prediction model. The blue path, on the other hand, delineates the ground truth—the actual movement of the worker following the last known point of the input path. This serves as the benchmark against which the model's predictive accuracy is measured. The model's prediction is visualized through black dashed lines, extending from the end of the input path to forecast the worker's future movements.

From Table 1 the LSTM model for path prediction has ADE values ranging from 0.41 to 1.09 and FDE values between 0.88 and 2.41. It performs best on the ZARA1 dataset, indicating its efficacy in less dynamic environments but struggles with the more complex Worker dataset, suggesting limitations in handling intricate social interactions and environmental complexities.

STGCNN significantly improves upon LSTM, especially in the HOTEL and ZARA2 datasets, showcasing its strength in capturing spatial-temporal relationships. Its lowest ADE and FDE values in the ZARA2 dataset highlight its capability to accurately predict paths in highly interactive environments.

STGAT demonstrates a marginal improvement over STGCNN in the HOTEL dataset but exhibits a slight increase in error rates in the Worker dataset. This suggests that while the attention mechanism provides an edge in certain scenarios, it may not always translate to improved performance in complex settings STGCNN v2 shows an interesting balance, with slight improvements over the original STGCNN in most datasets except for a marginal increase in ADE and FDE in the ETH and Worker datasets. This indicates that version 2 enhancements primarily benefit spatial-temporal reasoning without significantly impacting the model's overall predictive capabilities.

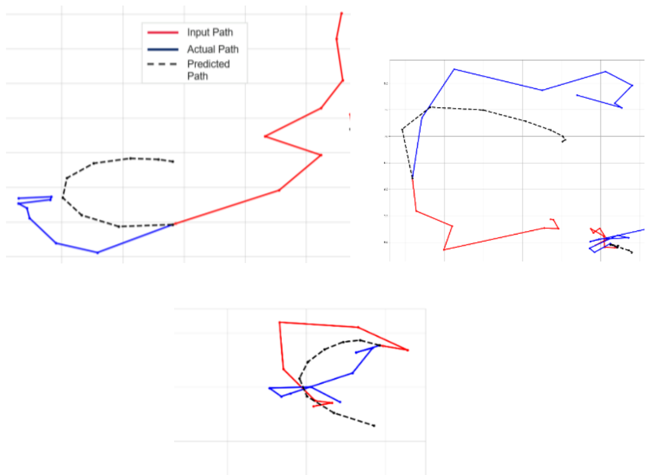


FIGURE 1: Trajectory Prediction: A Comparative Visualization of Input Path (Red), Ground Truth (Blue), and Predicted Path (Black Dashed Lines)

STGAT v2 offers the best performance among all models, particularly in the HOTEL and ZARA1 datasets. The updated attention mechanisms and model refinements contribute to its superior ability to navigate complex social interactions, although it still faces challenges in the Worker dataset, like its predecessors.

The variations in evaluation metrics between the models highlight the intricate balance between capturing spatial-temporal dynamics and accurately modeling social interactions. While LSTM serves as a strong baseline, the introduction of graph convolutional networks (STGCNN) and attention mechanisms (STGAT) markedly improve performance, especially in scenarios with rich pedestrian interactions.

The second versions of STGCNN and STGAT (v2) illustrate incremental improvements, with STGAT v2 leading in terms of both ADE and FDE across most scenarios. This suggests that while foundational model architectures provide substantial benefits, iterative refinements, especially in attention mechanisms, offer critical enhancements in prediction accuracy.

The comparative analysis underscores the importance of model choice based on the specific characteristics of the environment and the nature of human interactions within it. While STGCNN models excel in environments with clear spatial-temporal patterns, STGAT models, with their attention mechanisms, are better suited for scenarios where individual interactions significantly influence movement patterns.

The consistent challenge across models in accurately predicting worker movements in the Worker dataset indicates a need for further research into models that can more adeptly handle the unpredictability and complexity of industrial environments. This suggests an opportunity for integrating additional contextual information, such as environmental constraints and individual worker objectives, to further refine prediction accuracy.

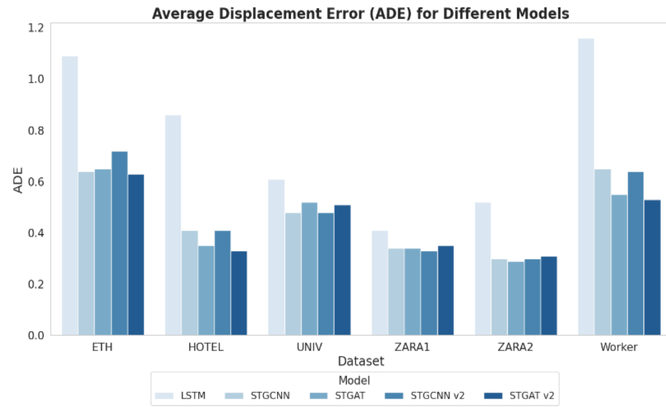


FIGURE 2: Comparative Analysis of Average Displacement Error (ADE) in meters Across Different Models and Datasets in Worker Trajectory Prediction.

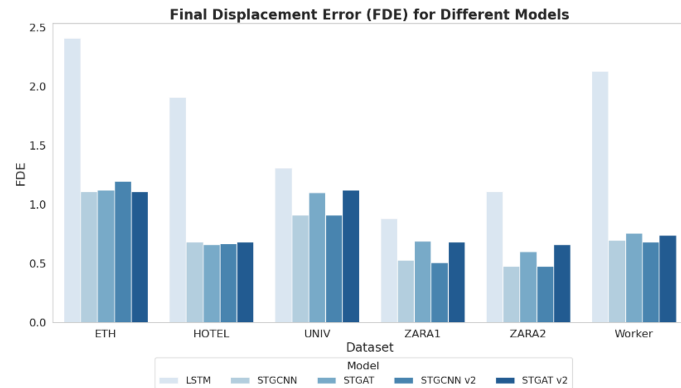


FIGURE 3: Comparative Analysis of Final Displacement Error (FDE) in meters Across Different Models and Datasets in Worker Trajectory Prediction.

4.3 Qualitative Analysis

This qualitative analysis explores how the model's attention weights adapt to different spatial and behavioural dynamics among workers, such as dwelling, transitioning, and proximity variations and while exchanging assembled sub parts.

Scenario 1: All Workers Dwelling

Description: In this scenario, all workers are stationary, focusing on tasks at their respective stations without moving significant distances.

Attention Mechanism Implications:

- The attention weights in this scenario might be lower between distant workers, as their current stationary states have minimal influence on each other's future movements.
- Nearby workers could have moderate attention weights $(0.4 < \alpha_{ij}^{t} < 0.7)$, reflecting the potential for interaction should one decide to move. However, the overall attention levels might remain subdued due to the general lack of movement.

In the visualizations shown below in Figure 4 depicting attention weights, each image focuses on one worker (T), represented without circles, while other workers are denoted by differently colored circles denoting different time steps. The size of the circles varies across different time steps, indicating the magnitude of attention weights assigned to each worker with respect to the target.

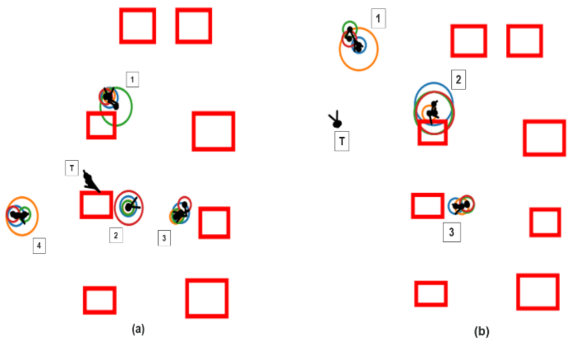


FIGURE 4: Attention weights visualized for Workers Dwelling and stationary at their position for t = *8 sec.*

Scenario 2: N-Workers in Transition, Others Dwelling

Description: Among a group of workers, two are moving between locations (in transition) while the others remain stationary at their workstations as shown in Figure 5.

Attention Mechanism Implications:

- The transitioning workers are likely to have higher attention weights between themselves, especially if their paths are converging or if they are moving within close proximity, indicating a higher level of interaction and potential collision avoidance.
- Stationary workers near the paths of the transitioning workers might exhibit increased attention weights towards these moving workers, reflecting the model's anticipation of possible interactions or obstructions.
- Distant dwelling workers may still have minimal attention weights towards the transitioning workers, given the low immediate impact on their movements.

Scenario 3: Workers Working Nearby Each Other

Description: Workers are in close proximity to each other, either stationary or performing tasks that require minimal movement as depicted in Figure 6.

Attention Mechanism Implications:

- High attention weights ($0.7 < \alpha_{ij}^{t} < 1.0$) are expected among these workers, as their close proximity increases the potential for interaction. This includes sharing tools, materials, or even navigating around each other.
  The model's attention mechanism would prioritize the spatial dynamics within this cluster of workers, predicting their movements with consideration to maintaining safe distances and efficient task collaboration.

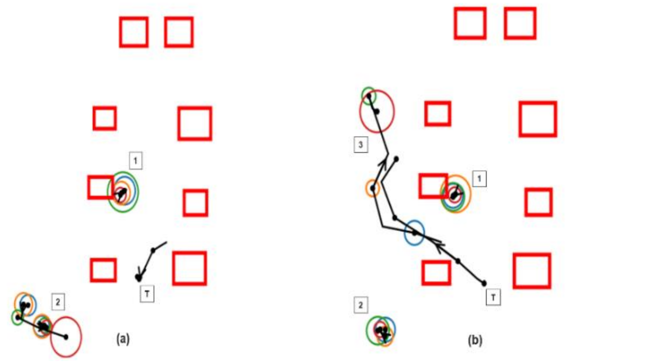
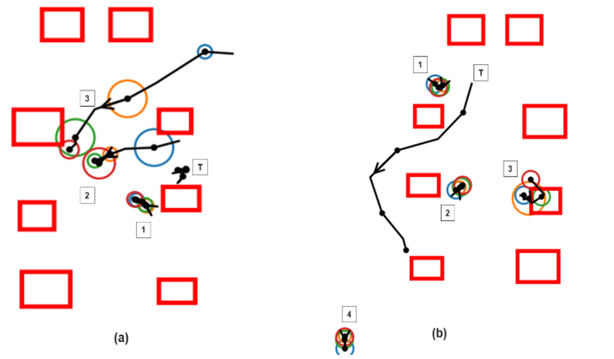


FIGURE 5: (a) Attention weights visualized for One Worker (2) in Transit and others stationary at their position **(b)** Two workers (Target and Worker 3) in transit and other workers dwelling for t



= *8 sec.*

FIGURE 6: **(a)** Scenario of two workers in transition and Target worker dwelling **(b)** Target worker in transition and other workers dwelling.

- The model's attention mechanism would prioritize the spatial dynamics within this cluster of workers, predicting their movements with consideration to maintaining safe distances and efficient task collaboration.

In Figure 7 (a) The Target worker T works nearby Worker 1 and hence a high attention weights are associated indicating high potential for interaction. Whereas the other workers away were given a lower and similar attention weights as they do not interact with the target worker and are static at their workplaces. Similarly in Figure 7 (b) Target worker and worker 1 are working near each other indicating higher attention weights associated to worker 1.

Scenario 4: Workers Far Apart from Each Other

Description: Workers are dispersed across a wide area, with significant distances separating them as shown in Figure 8.

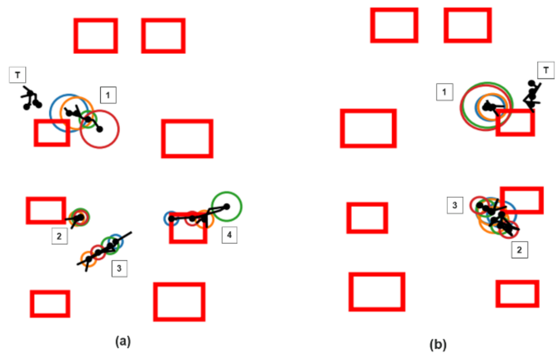


FIGURE 7: Workers working nearby each other for t = 8 secs

Attention Mechanism Implications:

- Low attention weights ($0 < \alpha_{ij}^{t} < 0.4$) would be characteristic of this scenario, as the spatial separation minimizes direct interactions. Each worker's movement is less likely to influence the immediate actions of others.
- The model may focus more on the individual movement patterns and tasks of each worker, rather than the complex inter-worker dynamics present in more crowded settings.

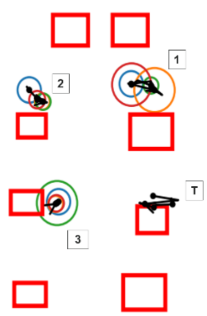


FIGURE 8: Visualization of attention weights for workers Far apart from each other.

4.3 Limitations and Drawbacks

Regarding the limitations and drawbacks of our approach in predicting worker trajectories within manufacturing settings, several aspects of our study merit further exploration and critique. Firstly, while our methodology significantly advances the precision of trajectory predictions using Spatio-temporal Graph Attention Network (STGAT) and Social-STGCNN models, it is primarily reliant on the accuracy of position data collected via sensors. Sensor inaccuracies or the occurrence of workers performing tasks outside predefined zones could lead to discrepancies in activity event detection. This is a critical consideration since our detection mechanism hinges on spatial data and predefined factory layout, marking regions assigned for specific tasks.

While the Average Displacement Error (ADE)/ Final Displacement Error (FDE) metrics are commonly utilized for evaluating trajectory prediction models, they present an incomplete picture of a model's prediction quality and performance. The Best-of-N (BoN) metric, though attempting to address this, does not quantify all generated samples, thus not fully capturing the model's predictive capability. To overcome this, introducing metrics like Average Mahalanobis Distance (AMD) and Average Maximum Eigenvalue (AMV) could provide a more comprehensive understanding of how closely predictions align with actual trajectories and the overall spread of predictions [45].

The accuracy of trajectory predictions in industrial settings can be significantly impacted by unforeseen events, varying machine states, and dynamic factory layouts. However, the current dataset utilized in our research lacks representation of such factors. Worker trajectories are influenced by a multitude of variables including production planning, machine faults, time of day, date of the year, accidents, and individual worker characteristics. The absence of these diverse examples and related input variables in the dataset poses a limitation to our study, as our models are unable to account for these real-world scenarios during prediction, potentially leading to errors. Nevertheless, as the dataset size grows to encompass a wider range of industrial operating conditions, these limitations can be mitigated. Furthermore, the dataset used in our research lacks labels corresponding to specific events, thereby limiting the depth of qualitative analysis that can be conducted. The absence of event-specific labels hinders our ability to verify qualitative insights into worker behaviours and interactions. Addressing these limitations would enhance the robustness and applicability of our methodology in real-world industrial environments.

4.4 Future Work

The future work stemming from this research opens several avenues for exploration and refinement. An immediate direction involves enhancing the accuracy of task-specific worker movement prediction by integrating more sophisticated sensor technologies and data fusion techniques. Additionally, expanding the dataset to include labelled events related to specific manufacturing tasks will allow for more granular analysis and validation of the proposed models against ground truth data.

To provide a comprehensive understanding of the proposed methodology's efficacy, a thorough comparison with existing approaches should be done. This will involve outlining the performance metrics and highlighting the distinctive features of the proposed STGAT and STGCNN models. Such comparisons are crucial for elucidating the relative strengths and weaknesses of these models. Testing the dataset on some of the current state-of-the-art models is necessary to further validate our findings and establish a solid benchmark.

Most of the state-of-the-art models have better evaluation metrics scores on benchmark datasets when compared to the above models but lack the internal decision-making evaluation essential to observe the attention metric while predicting the path. To enhance the paper's contribution, further evaluation using larger and more diverse datasets from various industrial settings is warranted to assess the robustness and applicability of the proposed approach. Incorporating newly proposed evaluation metrics such as Average Mahalanobis Distance (AMD) and Average Maximum Eigenvalue (AMV) [45] across more diverse datasets will further validate their effectiveness and reliability over traditional metrics like ADE/FDE.

5. CONCLUSION

In conclusion, this study presents a pioneering approach to worker trajectory prediction within manufacturing environments using state-of-the-art Spatio Temporal Graph Attention Network models. Our research demonstrates the potential of leveraging advanced machine learning techniques to predict worker movements and interactions, offering significant contributions to enhancing factory safety and operational efficiency. By integrating graph-based models and attention mechanisms, we have developed a method capable of capturing the complex dynamics of worker movements and providing insights into worker interactions through qualitative analysis of attention weights. This research underscores the importance of adopting advanced predictive analytics in the manufacturing industry. As factories become increasingly integrated and automated, the ability to accurately model and predict worker movements will play a critical role in ensuring safety, minimizing operational disruptions, and optimizing production processes. Our work contributes to the growing body of knowledge in smart manufacturing, paving the way for more intelligent, responsive, and efficient manufacturing systems.

REFERENCES

[1] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, “Human Motion Trajectory Prediction: A Survey Journal Title XX(X):1-37 c”, doi: 10.1177/ToBeAssigned.

[2] K. Lyu, H. Chen, Z. Liu, B. Zhang, and R. Wang, “3D human motion prediction: A survey,” 2022, doi: 10.1016/j.neucom.2022.02.045.

[3] X. Xia, T. Zhou, J. Du, and N. Li, “Human motion prediction for intelligent construction: A review,” 2022, doi: 10.1016/j.autcon.2022.104497.

[4] B. I. Sighencea, R. I. Stanciu, and C. D. Căleanu, “A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction,” *Italian National Conference on Sensors*, vol. 21, no. 22, Nov. 2021, doi: 10.3390/S21227543.

[5] M. Gulzar, Y. Muhammad, and N. Muhammad, “A Survey on Motion Prediction of Pedestrians and Vehicles for Autonomous Driving,” *IEEE Access*, vol. 9, pp. 137957–137969, 2021, doi: 10.1109/ACCESS.2021.3118224.

[6] R. Korbmacher and A. Tordeux, “Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-based Approaches”.

[7] J. Belter, M. Hering, and P. Weichbroth, “Motion Trajectory Prediction in Warehouse Management Systems: A Systematic Literature Review,” *Applied Sciences (Switzerland)*, vol. 13, no. 17, Sep. 2023, doi: 10.3390/APP13179780.

[8] M. M. Aliabadi, E. Darvishi, R. Shahidi, F. Ghasemi, and M. Mahdinia, “Explanation and prediction of accidents using the path analysis approach in industrial units: The effect of safety performance and climate.,” *Work*, vol. 66, no. 3, pp. 617–624, 2020,

[9] A. Löcklin, F. Dettinger, M. Artelt, N. Jazdi, and M. Weyrich, “Trajectory Prediction of Workers to Improve AGV and AMR Operation based on the Manufacturing Schedule,” *Procedia CIRP*, vol. 107, pp. 283–288, 2022, doi: 10.1016/j.procir.2022.04.046.

[10] A. Aslan *et al.*, “Using Worker Position Data for Human-Driven Decision Support in Labour-Intensive Manufacturing,” 2023, doi: 10.3390/s23104928.

[11] J. Cai, A. Du, X. Liang, and S. Li, “Prediction-Based Path Planning for Safe and Efficient Human-Robot Collaboration in Construction via Deep Reinforcement Learning,” *Journal of computing in civil engineering*, vol. 37, no. 1, Jan. 2023, doi: 10.1061/(ASCE)CP.1943-5487.0001056.

[12] D. Hu, S. Li, J. Cai, and Y. Hu, “Toward Intelligent Workplace: Prediction-Enabled Proactive Planning for Human-Robot Coexistence on Unstructured Construction Sites,” *Online World Conference on Soft Computing in Industrial Applications*, vol. 2020-December, pp. 2412–2423, Dec. 2020, doi: 10.1109/WSC48552.2020.9384077.

[13] X. Gong, T. Wang, Y. Cui, T. Huang, T. Wang, and Y. Cui, Towards Safe and Efficient Swarm-Human Collaboration: A Hierarchical Multi-Agent Pickup and Delivery framework, *IEEE Transactions On Intelligent Vehicles*, 1, 2022.

[14] S. Hong, K. Jang, J. Lee, H. Yoon, HyungtaeCho, and I. Moon, “Optimal Evacuation Route Prediction in Fpso Based on Deep Q-Network,” *Computer Aided Chemical Engineering*, vol. 48, pp. 1867–1872, Jan. 2020, doi: 10.1016/B978-0-12-823377-1.50312-8.

[15] S. Hong, K. Jang, J. Lee, H. Yoon, HyungtaeCho, and I. Moon, “Optimal Evacuation Route Prediction in Fpso Based on Deep Q-Network,” *Computer Aided Chemical Engineering*, vol. 48, pp. 1867–1872, Jan. 2020, doi: 10.1016/B978-0-12-823377-1.50312-8.

[16] D. Kim, S. Lee, and V. R. Kamat, “Proximity Prediction of Mobile Objects to Prevent Contact-Driven Accidents in Co-Robotic Construction,” *Journal of Computing in Civil Engineering*, vol. 34, no. 4, Jul. 2020, doi: 10.1061/(ASCE)CP.1943-5487.0000899.

[17] L. Luo and N. Zhao, “An efficient simulation model for layout and mode performance evaluation of robotic

mobile fulfillment systems," *Expert Syst Appl*, vol. 203, p. 117492, 2022, doi: 10.1016/j.eswa.2022.117492.
[18] H. Li, X. Yang, F. Wang, T. Rose, G. Chan, and S. Dong, "Stochastic state sequence model to predict construction site safety states through Real-Time Location Systems," 2015, doi: 10.1016/j.ssci.2015.11.025.
[19] Tang, S., Golparvar-Fard, M., Naphade, M., & Gopalakrishna, M. M. (2020). Video-based motion trajectory forecasting method for proactive construction safety monitoring systems. *Journal of Computing in Civil Engineering*, 34(6), 04020041.
[20] X. Tao and W. Song, "Profit-Oriented Task Allocation for Mobile Crowdsensing with Worker Dynamics: Cooperative Offline Solution and Predictive Online Solution," *IEEE Trans Mob Comput*, vol. 20, no. 8, pp. 2637–2653, Aug. 2021, doi: 10.1109/TMC.2020.2983688.
[21] Helbing, D., and Molnar, P. (1995). Social force model for pedestrian dynamics. Physical review E, 51(5), 4282.
[22] J. Liang, L. Jiang, J. C. Niebles, A. Hauptmann, and L. Fei-Fei, "Peeking into the Future: Predicting Future Person Activities and Locations in Videos", Accessed: Mar. 11, 2024. [Online]. Available: https://next.cs.cmu.edu
[23] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, "SR-LSTM: State Refinement for LSTM towards Pedestrian Trajectory Prediction".
[24] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks".
[25] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, S. H. Rezatofighi, and S. Savarese, "SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints".
[26] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, S. H. Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal Trajectory Forecasting using Bicycle-GAN and Graph Attention Networks".
[27] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction", Accessed: Mar. 04, 2024. [Online]. Available: https://github.
[28] R. Liang, Y. Li, J. Zhou, and X. Li, "STGlow: A Flow-Based Generative Framework With Dual-Graphormer for Pedestrian Trajectory Prediction," *IEEE Trans Neural Netw Learn Syst*, 2023, doi: 10.1109/TNNLS.2023.3294998.
[29] A. Vemula, K. Muelling, and J. Oh, "Social Attention: Modeling Attention in Human Crowds".
[30] F. Giuliari, I. Hasan, M. Cristani, and F. Galasso, "Transformer Networks for Trajectory Forecasting".
[31] K. Mangalam *et al.*, "It is not the Journey but the Destination: Endpoint Conditioned Trajectory Prediction", Accessed: Mar. 08, 2024. [Online]. Available: https://karttikeya.github.io/publication/htf/
[32] R. C. Mercurius, E. Ahmadi, S. Mohamad, A. Shabestary, and A. Rasouli, "AMEND: A Mixture of Experts Framework for Long-tailed Trajectory Prediction".
[33] C. Xu, W. Mao, W. Zhang, and S. Chen, "Remember Intentions: Retrospective-Memory-based Trajectory Prediction", Accessed: Mar. 08, 2024. [Online]. Available: https://github.com/MediaBrain
[34] H. Zhao and R. P. Wildes, "Where are you heading? Dynamic Trajectory Prediction with Expert Goal Examples", Accessed: Mar. 08, 2024. [Online]. Available: https://github.com/JoeHEZHAO/expert_traj
[35] G. Alinezhad Noghre, V. Katariya, A. Danesh Pazho, C. Neff, and H. Tabkhi, "Pishgu: Universal Path Prediction Network Architecture for Real-time Cyber-physical Edge Systems; Pishgu: Universal Path Prediction Network Architecture for Real-time Cyber-physical Edge Systems", doi: 10.1145/3576841.3585933.
[36] J. Yue, D. Manocha, and H. Wang, "Human Trajectory Prediction via Neural Social Physics", Accessed: Mar. 08, 2024. [Online]. Available: https://github.com/realcrane/Human-
[37] K. Mangalam, Y. An, H. Girase, and J. Malik, "From Goals, Waypoints & Paths To Long Term Human Trajectory Forecasting," Dec. 2020, [Online]. Available: http://arxiv.org/abs/2012.01526
[38] C. Wong *et al.*, "View Vertically: A Hierarchical Network for Trajectory Prediction via Fourier Spectrums", Accessed: Mar. 08, 2024. [Online]. Available: https://github.com/cocoon2wong/Vertical.
[39] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction".
[40] S. Brody, U. Alon, and E. Yahav, "How Attentive Are Graph Attention Networks?", Accessed: Mar. 04, 2024. [Online]. Available: https://github.com/tech-srl/how_attentive_are_
[41] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces".
[42] P. Veličkovi´veličkovi´c, G. Cucurull, A. Casanova, A. Romero, P. Lì, and Y. Bengio, "GRAPH ATTENTION NETWORKS".
[43] M. Delamare, F. Duval, and R. Boutteau, "A New Dataset of People Flow in an Industrial Site with UWB and Motion Capture Systems", doi: 10.3390/s20164511.
[44] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Computer Graphics Forum*, vol. 26, no. 3, pp. 655–664, 2007, doi: 10.1111/J.1467-8659.2007.01089.X.
[45] A. Mohamed, D. Zhu, W. Vu, M. Elhoseiny, and C. Claudel, "Social-Implicit: Rethinking Trajectory Prediction Evaluation and The Effectiveness of Implicit Maximum Likelihood Estimation", Accessed: Mar. 08, 2024.

