IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 May 2024, accepted 22 August 2024, date of publication 28 August 2024, date of current version 17 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3450955

RESEARCH ARTICLE

WTTFNet: A Weather-Time-Trajectory Fusion Network for Pedestrian Trajectory Prediction in Urban Complex

HO CHUN WU[1], (Member, IEEE), ESTHER HOI SHAN LAU[2], PAUL CHUN HO YUEN[1], KEVIN HUNG[1], (Senior Member, IEEE), JOHN KWOK TAI CHUI[1], (Member, IEEE), AND ANDREW KWOK FAI LUI[1]

[1]School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, China
[2]School of Nursing and Health Studies, Hong Kong Metropolitan University, Hong Kong, China

Corresponding author: Kevin Hung (khung@hkmu.edu.hk)

This work was supported by the Research Grants Council of Hong Kong Special Administrative Region, China, under Grant UGC/FDS16/E12/20.

ABSTRACT Pedestrian trajectory modelling in an urban complex is challenging because pedestrians can have many possible destinations, such as shops, escalators, and attractions. Moreover, weather and time-of-day may affect pedestrian behavior. In this paper, a new weather-time-trajectory fusion network (WTTFNet) is proposed to incorporate weather and time-of-day (WT) information to refine the predicted destination and trajectories. First, a word embedding is used to encode the WT information and its representation can be further optimized according to the loss function. Afterwards, a gate multimodal unit is used to fuse the WT information and preliminary pedestrian intent probabilities obtained from a preliminary baseline classifier. A joint loss function based on focal loss is used to co-optimize both the preliminary and final classifiers, which helps to improve the accuracy under possible class imbalances. Finally, a destination adapted trajectory model is used predict the trajectories guided by the predicted destination. Experimental results using the Osaka Asia and Pacific Trade Center (ATC) dataset shows improved performance of the proposed approach over state-of-the-art algorithms by 23.67% increase in classification accuracy, 9.16% and 7.07% reduction of average and final displacement error. The proposed approach may serve as an attractive approach for improving existing baseline trajectory prediction models when they are applied to scenarios with influences of weather-time conditions. It can be employed in numerous applications such as pedestrian facility engineering, public space development and technology-driven retail.

INDEX TERMS Functional objects, LSTM, pedestrian trajectory prediction, urban complex, weather.

I. INTRODUCTION

Predicting pedestrian trajectories in crowd scenario is essential in smart city. It has numerous applications such as self-driving cars [1], smart road crossings and intelligent retail [2]. KB models describe pedestrian dynamics using physical, social or psychological rules. Pioneer KB models are Social Force model [3] and collision avoidance [4]. Deep learning (DL) approaches leverage extensive observations. They can be mainly categorized into Recurrent Neural Network (RNN) [5], [6], Convolutional Neural Network (CNN) [7], Transformer (TF) [8], Generative Adversarial Network (GAN) [9], [10], [11]. Most recent research focuses on Social-awareness incorporated deep neural network architectures [10] and graph convolutional network (GCN) [12] to further improve performance.

While much attention is directed towards modelling the trajectory in outdoor scenarios with applications to autonomous vehicles, this paper focuses on modelling the pedestrian trajectory within an urban complex. Recently, Indoor Pedestrian Trajectory Generator (IPTG) [13] was reported, which uses a GAN based approach to generate trajectories for a fictional

The associate editor coordinating the review of this manuscript and approving it for publication was Ángel F. García-Fernández.

conference scenario. Han et al. [14] employed trajectory clustering in modeling pedestrian flow for indoor design space. D'Orazio et al. [15] simulated the pedestrian flow of a building using agent-based model with proximity and exposure time based rules to estimate the spread of Coronavirus Disease (COVID-19) in building. However, there is few existing literature about indoor pedestrian trajectory modelling under the influence of weather and time-of-day, a.k.a. weather-time (WT) condition. Xue et al. [16] studied the modelling of pedestrian movement in a train station and proposed a Pedestrian Trajectory Prediction method by LSTM with Automatic Route Class Clustering (PoPPL). It employed k-mean clustering to label pedestrian trajectories followed with subsequent LSTM based intent classification and trajectory prediction. However, the train station dataset only contained video lasting for 30 minutes with same weather and it mainly serves the purpose of transportation.

Weather-time (WT) conditions refer to weather and time-of-day variations. An objective of this paper is to study the effect of weather and time-of-day for pedestrian movement pattern in urban complex. Typical indoor environment, such as residential apartments, offices, factories, etc., are single functional premises. Individuals usually share common location-of-interest (LOI), i.e. going home/going to work. In contrast, pedestrian behavior in urban complexes exhibits much more randomness as the pedestrians could have different destinations to functional objects [17] that serves a wide range of purposes, such as retail, shopping malls, office accommodations, and business functions. Previous studies [18], [19] suggested that weather has an impact in affecting pedestrian behavior. In particular, bad weather may discourage consumers from shopping. Also, adverse weather conditions may lead to delays or cancellations of public transportation services [20], which affects pedestrian traffic. Time-of-day will affect commuter traffic and hence pedestrian flow [21], [22]. This study aims to improve understanding on how the weather and time-of-day influence the choice of destination and hence the trajectories of pedestrians, which will help to facilitate flow management [23] and intelligent retail [2]. With the increasing popularity of multimodal transportation in large metropolises to decrease reliance on private cars and greenhouse emission, many urban complexes are designed with multimodal transportation [24] capabilities. They serve as interconnection points to facilitate seamless transfers between buses and trains. Examples are Osaka station (Osaka, Japan) [24] and Chatswood interchange shopping mall (Sydney, Australia).

Three practical issues may arise in modelling the pedestrian trajectory under different weather-time (WT) conditions in urban complex are i) appropriate preprocessing and feature selection, ii) effective fusion, iii) choice of clusters under the effect of different WT conditions.

First, the format of weather information may not directly fit for use and require appropriate preprocessing and feature selection. Directly concatenating this information to the deep neural network may even confuse the classifier and lead to inferior performance. For instance, Time-of-day information is commonly available as numeric values and the classifier may perceive it as ordinal, i.e. 9 o'clock is larger than 8 o'clock, which is not logical at all.

Second, it is not trivial on where and how to fuse the WT information. For example, direct concatenation of one-hot encoded WT information to the raw pedestrian trajectories does not yield satisfactory performance.

Third, although the use of trajectory prediction guided by pedestrian intent have been reported before, it is mainly used to predict the pedestrian's intent for road crossing in outdoor scenarios [25] involving pedestrian-vehicle interaction. Unlike the road crossing scenario, where pedestrians will need to cross the road under different weather conditions, the pedestrian behavior in urban complex can be affected by weather, especially in destinations for retail and entertainment.

To overcome these challenges in improving the pedestrian trajectory prediction accuracies of baseline deep learning models, we propose a new weather-time-trajectory network (WTTFNet) for pedestrian trajectory prediction The WTTFNet is made up of the following components:

1. Weather-time (WT) Embedding: To tackle the issue of preprocessing and feature selection of WT information, a word embedding is used to encode the WT information and it has the advantage to be further optimized according to the final loss function.
2. A new statistical test based on the Pearson's chi-squared χ^2 statistic is used to test the significance of the WT condition and determine whether to incorporate the WT information.
3. Novel WTTFNet based intended destination (ID) classifier: The ID classifier is used to predict the destination based on input trajectories. Motivated by the rationale that weather-time conditions can influence the decision of reaching a destination, the proposed WTTF architecture employs the Gated Multimodal Unit (GMU) to fuse the WT embedding with preliminary pedestrian intent probabilities obtained from a baseline deep neural network based classifier. The fused representation is used to train the final classifier, which generates predicted destinations refined by the weather and time.
4. Deep supervision [27] is used to co-train the preliminary and final classifiers together using auxiliary and final loss functions. While the preliminary pedestrian intent probabilities provide supervisory signals to train the baseline classifier, the final loss function optimizes the whole architecture. The Focal Loss [28] is used to cater for possible class imbalance. A Destination adapted trajectory predictor (DATP) is used to perform subsequent trajectory prediction. Multiple trajectory models targeted to different destinations are trained and the trajectory model that points to the predicted destination will be chosen.

To illustrate the effectiveness of the proposed approach in improving a baseline pedestrian trajectory model, the public dataset obtained from Asia and Pacific Trade Center (ATC) [29] in Osaka is considered. It is an urban complex serving as

a multimodal transportation hub, which connects the intercity ferry pier and Osaka metro line, as well as accommodating a trade center and multi-entertainment complex. Pedestrian trajectories obtained on a sunny (22nd May, 2013) and cloudy day (29th September, 2013) were used. There were roughly 1.5 times more pedestrians during peak hours in compared to off-peak hours. A significant log p-value of -104.8395 ($\ll \log(0.05)$))[1] is attained using the proposed statistical test, which suggests that there is significant deviation in pedestrian flow across weather and off/peak hours.

Experimental results show that the proposed WTTFNet surpasses state-of-the-art algorithm by reduction of 9.16 % and 7.07% in average displacement error (ADE) and final displacement error (FDE), respectively. It also improves the classification accuracy (ACC) and Cohen's Kappa (κ) of the baseline model (i.e. PoPPL) by 23.67% and 28.13%, respectively.

To study the role of weather and time-of-day in improving prediction performance, ablation test is performed to compare between the proposed WTTFNet with/without incorporation of weather-time information. Significant McNemar's test [30] p-value of $p = 0.0196 < 0.05$ was attained, which suggests the improvement in classification accuracy from 71.5% to 71.95% after adding weather-time information was significant because of the large sample size of 28536 pedestrians.

Further analysis of the 3008 significant pedestrians identified by McNemar's test shows that an overall 5.47% (7.8m to 7.4m) and 7.58% (14.11m to 13.04m) improvement in ADE and FDE reduction were obtained for the significant 3008 pedestrians, and significant one-sided Mann–Whitney U test [32] p-values were attained for ADE ($p = 0.0203 < 0.05$) and FDE ($p = 0.00533 < 0.05$), respectively. This shows that weather-time information helps to improve prediction performance significantly for the 3008 cases considered. Overall, the ratio 3008 out of 28536 pedestrians was also statistically significant according to the McNemar's test, suggesting that these 3008 pedestrians showing significantly improved performance out of 28536 cases were very unlikely a random event. This suggests the proposed approach may serve as an attractive approach for incorporating WT information to improve pedestrian trajectory prediction and it also serves as a systematic approach to test the significance of WT conditions.

Finally, with the increasing popularity of multimodal transportation in large metropolises to decrease reliance on private cars and reduce greenhouse emission, understanding pedestrians' behavior in urban complex is increasingly important. Walking networks with interconnecting urban complexes will be increasingly prevalent to facilitate smooth transfers between different modes of transportation and contribute to the economic development of nearby areas. There are also numerous applications in public space development [33], evacuation planning [34], and advancements in technology-driven retail [2].

[1] A significance level of 0.05 is sought [31].

The rest of this paper is organized as follows. Section II presents a review on the background and related works, whereas the proposed WTTFNet is presented in Section III. In Section IV, experimental results and comparisons with state-of-the-art algorithms are presented. The proposed statistical test is also used to test the significance of weather-time effects. Finally, conclusion is drawn in Section V.

II. BACKGROUND AND RELATED WORK

Pedestrian trajectory prediction (PTP) methods can be categorized according to input modality, network architecture, features, and prediction tasks [35], [36]. Traditionally, PTP is achieved using knowledge based methods such as social force [3] collision avoidance [4], kinetic models [37]. In the last decade, deep learning approaches have gained much popularity for its powerfulness in leveraging extensive observations. They can be mainly categorized into

1. Recurrent neural network (RNN): Examples are Long Short Term Memory (LSTM) [5], Social LSTM [38], Gated Recurrent Unit (GRU) and Conv-LSTM [39]. LSTM are renowned for its capability to handle sequence-to-sequence prediction. Social LSTM further extends LSTM to model social interactions. Conv-LSTM replaces the fully connected layers in conventional LSTM with convolutional layers, which enables the capturing of both spatial and temporal information for intent and trajectory prediction in [39].
2. Convolutional neural networks (CNN): The CNNs are usually used for PTP approaches that uses images/videos to predict the trajectories. CNN is used to extract spatial-temporal features [7] or skeleton keypoints [40] for classifying pedestrian behaviour.
3. Transformer: VOSTN [8] used a variational one-shot transformer for trajectory prediction together with a cross-attention module to model the inter-relationship between trajectory and ego-motion. AgentFormer [10] integrated a transformer architecture with agent-aware attention mechanism and a conditional variational autoencoders (CVAE) based trajectory prediction framework.
4. Generative adversarial network (GAN): POI-GAN [41] used generative model that integrates interest point model, field of view angle, and observed trajectories, to produce projected pedestrian trajectories for future time frames. Social GAN [9] employs a LSTM model to capture temporal structure of individual pedestrian and a social pooling mechanism to aggregate pedestrian interactions. The resultant deep features are used to train the GAN.

Over the past 5 years, most research focuses on incorporation of Social-awareness [9], [10], or contextual information [25], [39] to improve prediction performance. Social-awareness approaches such as social LSTM Mann and Whitney [32], social GAN [9], Sophie [42], AgentFormer [10] etc., primarily center around predicting trajectories and modeling interactions among a fixed number of pedestrians based on social pooling mechanisms. GCN based

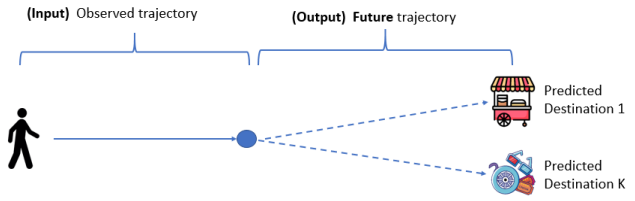


FIGURE 1. A pedestrian trajectory prediction problem. The observed trajectory is used to predict the future trajectory and final destination in this paper.

approaches, such as Social Spatial Temporal Graph CNN (SSTGCNN) [43], which models pedestrian interactions as graphs and extract spatial-temporal feature from the graphs using convolutional operations.

Context-based approaches incorporates context information to predict pedestrian intent and use it to guide subsequent trajectory prediction [23], [33]. Typical pedestrian intent includes crossing road and other walking gestures [44]. These intents are predicted from video or LIDAR sequences. Examples of contextual information are road topology, maps, pedestrian attributes, road boundaries and ego-vehicle information [23], [33].

While much attention is directed towards modelling the trajectory in outdoor scenarios with applications to autonomous vehicles, this paper focuses on modelling the pedestrian trajectory within an urban complex, which is challenging because pedestrians can have many possible destinations, such as shops, escalators, and attractions. Moreover, weather and time-of-day may affect pedestrian behavior. A new weather-time-trajectory fusion network (WTTFNet) is proposed to incorporate weather and time-of-day (WT) information to refine the predicted destination and trajectories. In the next section, the proposed methodology will be discussed.

III. PROPOSED METHODOLOGY

Fig. 1 shows an illustration of the pedestrian trajectory prediction problem, where the proposed WTTFNet predicts the final destination and future trajectory from partially observed trajectory, e.g. half of the trajectory in this paper. The proposed WTTFNet is made up of the following components:

1. Destination-driven clustering: It is used to label the pedestrian trajectories of the training set with destinations assigned by k-mean clustering for subsequent training of the intended-destination (ID) classifier.
2. The proposed statistical test based on the Pearson's chi-squared χ^2 statistic is designed to determine the minimum sample size required for each cluster and determine whether to incorporate the WT information.
3. ID classifier: It predicts the final destination that occurs in future from an observed "historical" trajectory of the pedestrian. The training set is provided by the destination-driven clustering. It is made up of a baseline deep neural network based classifier and the proposed WTTFNet, which serve as the preliminary and final classifiers, respectively. The baseline classifier will generate a set of preliminary pedestrian intent probabilities indicating the chances of reaching different destinations. Afterwards, the WTTFNet fuses the WT information and the preliminary pedestrian intent probabilities for subsequent training of the final classifier, which generates the final intent probabilities.
4. Destination adapted trajectory predictor (DATP): After the final pedestrian intent probabilities are generated, the destination with the highest probability is chosen. The target trajectory model trained using the clustered trajectories of surrounding the chosen destination is used to predict the future trajectory. As an illustration, the PoPPL-def sub-LSTM [16] is adopted as the trajectory model. In general, other trajectory prediction models can be used.

A. DESTINATION-DRIVEN CLUSTERING MODULE

In a pedestrian trajectory prediction problem, an observed trajectory $\boldsymbol{s_n}$ for the $n-th$ pedestrian of length L is used to predict the future L' observations trajectory $\hat{z}_{\boldsymbol{n}}$:

$$\boldsymbol{s_n} = \left\{\left(x_{n,1}, y_{n,1}\right), \ldots, \left(x_{n,L}, y_{n,L}\right)\right\} \tag{1a}$$

$$\hat{z}_{\boldsymbol{n}} = \left\{\left(\hat{x}_{n,L+1}, \hat{y}_{n,L+1}\right), \ldots, \left(\hat{x}_{n,L+L'}, \hat{y}_{n,L+L'}\right)\right\}. \tag{1b}$$

However, there are multiple possible destinations of a pedestrian and hence a destination-driven clustering will be beneficial for training destination-specific trajectory models. In the destination-driven clustering module, the end-point of all trajectories, i.e. $\Omega_{end} : \{\left(x_{n,L+L'}, y_{n,L+L'}\right), n = 1, 2, \ldots, N\}$ from (1b) are passed to the k-means algorithm to form clusters. The membership of an endpoint (x, y) is sought by minimizing its distance from the centroids $\sum_{k=1}^{K}\sum_{(x,y)\in S_k} ||\,(x, y) - \left(\mu_{x,k}, \mu_{y,k}\right)||_2^2$, where Ω_k is the $k-th$ cluster and its centroid is updated as $\boldsymbol{\mu_k} = \left[\mu_{x,k}, \mu_{y,k}\right]^T = \frac{1}{|\Omega_k|}\sum_{(x,y)\in\Omega_k}(x, y)$. $|\Omega_k|$ is the number of elements in Ω_k. After assignment, each trajectory is labelled with the corresponding class from $\omega = 1, \ldots, K$ for sub-sequent training of the pedestrian intent classifier. The raw trajectories are cleaned and resampled so that the total duration of each trajectory is normalized to T_o.

The proposed approach also employs a statistical test to test the significance of each cluster and establish the minimum number of samples for each cluster (See Eqn. (12)). If a cluster is found to have insufficient number of samples, it can be merged to one of the clusters using an agglomerative clustering similarity measure, such as centroid linkage criterion

$$min.||\left(\mu_{x,k}, \mu_{y,k}\right) - \left(\mu_{x,k_i}, \mu_{y,k_i}\right)||_2^2, \tag{2}$$

where $\left(\mu_{x,k}, \mu_{y,k}\right)$ is the centroid of the cluster to be merged and $\left(\mu_{x,k_i}, \mu_{y,k_i}\right)$ are the remaining clusters. It is noted that other similarity measures can be employed. After the clusters have been computed, the training dataset for the ID classifier can be obtained as

$$\text{Trajectory:}\boldsymbol{s_n} = \left\{\left(x_{n,1}, y_{n,1}\right), \ldots, \left(x_{n,L}, y_{n,L}\right)\right\}, \tag{3a}$$

$$\text{Destination:}\omega_n = 1, \ldots, K, \tag{3b}$$

where K is the total number of destinations.

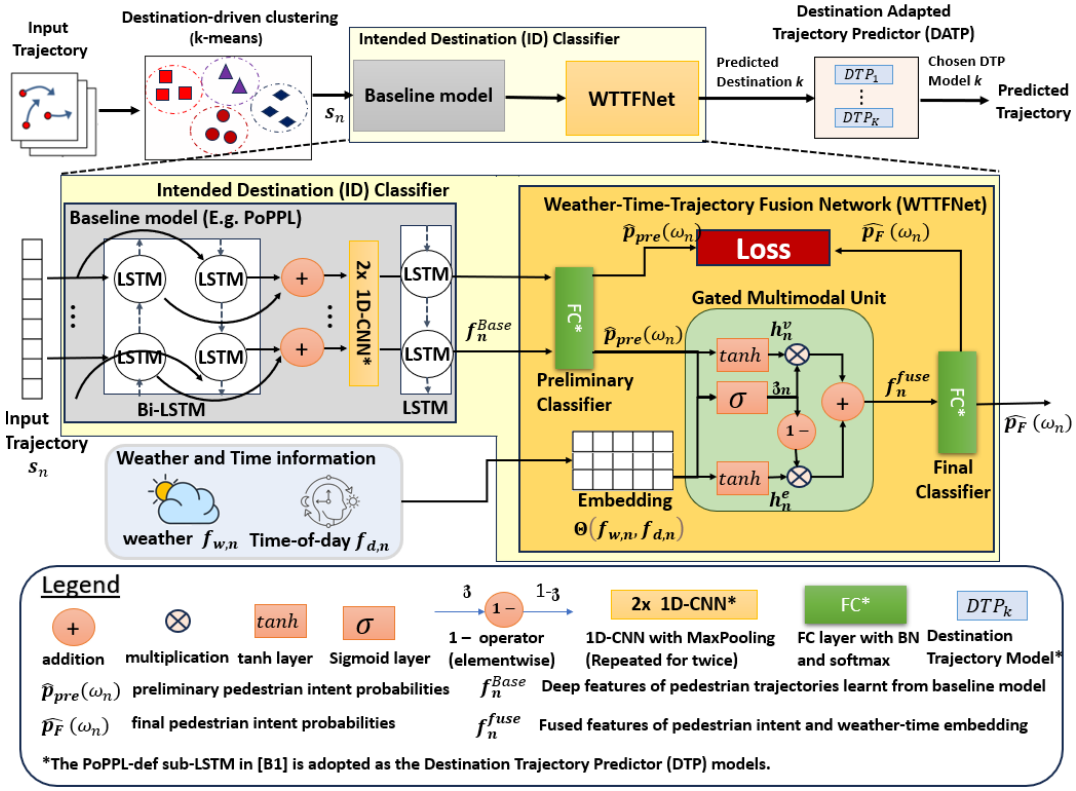


FIGURE 2. The proposed WTTFNet. Key innovations lie in the Intended Destination (ID) classifier, which is made up of i) baseline model, ii) focal loss iii) Deep supervision (preliminary and final classifiers optimized using joint loss function), and iv) incorporation of weather and time information via Gated Multimodal Unit (GMU). The structural details are summarized in Table 1.

B. NOVEL WEATHER-TIME-TRAJECTORY NETWORK FOR DESTINATION CLASSIFICATION

Fig. 2 and Table 1 show the proposed intended destination (ID) classifier, which comprises the weather-time (WT) embedding, baseline model (e.g. PoPPL) and the novel WTTFNet. First, a baseline model is used to learn the micro-level representation of the trajectory. Afterwards, a fully connected (FC) layer is used to learn a preliminary classifier of the destinations. The output preliminary ID class probabilities are then passed to the GMU for fusing with the WT embedding. The fused multimodal representation is passed to a final FC layer for training the final classifier. Both the preliminary and final classifier are co-optimized using the focal loss function. Here, the PoPPL is employed as the baseline model. In general, other trajectory models can be used.

More precisely, suppose there are C_w weather conditions and C_d different time-of-day and the total number of weather-time conditions are $C = C_w + C_d$. For example, in this paper, $C_w = 2$ (sunny/rainy) and $C_d = 2$ (off-peak/peak hours) are chosen. The proposed Weather-Time (WT) Embedding for the $n-th$ pedestrian is given as

$$\text{WT Embedding:} \boldsymbol{e_{WT,n}} = \boldsymbol{\Theta}\left(\boldsymbol{f_{w,n}}, \boldsymbol{f_{d,n}}\right), \tag{4}$$

where $\Theta()$ is the embedding layer. $\boldsymbol{f_{w,n}}$ and $\boldsymbol{f_{d,n}}$ are the one-hot encodings describing the weather-time condition for the $n-th$ pedestrian. The preliminary ID class probabilities $\hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n)$ can be obtained as the softmax probabilities from the preliminary classifier in Fig. 2. Batch normalization and softmax are performed after the FC layer. The preliminary intent probabilities $\hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n)$ and preliminary classifier $\boldsymbol{f_n^C}$ are given as

$$\hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n) = \boldsymbol{\sigma}_{\text{Soft}}\left(\boldsymbol{f_n^C}\right), \tag{5a}$$

$$\boldsymbol{f_n^C} = \boldsymbol{\phi_{BN}}\left(\text{FC}\left(\boldsymbol{f_n^{base}}\right)\right), \tag{5b}$$

respectively, where

$$\boldsymbol{\sigma}_{Soft}(u) = \frac{1}{\sum_{k=1}^{K} e^{u_k}}\left[e^{u_1}, e^{u_2}, \ldots, e^{u_K}\right]^T, \quad \text{and} \tag{5c}$$

$$\phi(u_k) = \frac{u_k - E(u_k)}{\sqrt{var(u_k) + \epsilon}} \times w_{\gamma,k} + w_{b,k} \tag{5d}$$

represent softmax operation and Batch normalization (BN), respectively. $\boldsymbol{f_n^C}$ and $\boldsymbol{f_n^{base}}$ are the output of the preliminary classifier and base model, respectively for the $n-th$ pedestrian. $\boldsymbol{\phi_{BN}}(\boldsymbol{u}) = [\phi(u_1), \phi(u_2), \ldots, \phi(u_K)]^T$ is the batch normalization function. $\text{FC}(\boldsymbol{u}) = \boldsymbol{W} \cdot \boldsymbol{u}$ is a fully connected layer with weights $\boldsymbol{W}$ and $\boldsymbol{\sigma}_{\text{Soft}}(\boldsymbol{u})$ is the softmax function. $w_{\gamma,k}$ and $w_{b,k}$ are learnable parameters for BN.

The preliminary pedestrian intent probabilities $\hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n)$ and the WT embedding $\boldsymbol{\Theta}\left(\boldsymbol{f_{w,n}}, \boldsymbol{f_{d,n}}\right)$ are then fused at

TABLE 1. Structural details of proposed WTTFNet.

Operation	Dimension (Input, Output)
(a) Preliminary Classifier (Eqn. (5b))	
Fully connected (FC) + Batch Normalization +Softmax	([a]K_{LSTM},K)
(b) Weather-Time (WT) Embedding (Eqn. (4))	
Embedding	([b]C,128)
(c) Gated Multimodal Unit (Eqns. (6a) to (6d))	
Tanh for Preliminary Classifier (Eqn. 6a)	(K,128)
Tanh for Embedding (Eqn. 6b)	(128,128)
Gate Neuron (Eqn. 6c)	(256,128)
1 – operator, Multiplication $\otimes$, and Addition $\oplus$	(128,128)
(d) Final Classifier (Eqn. (7))	
FC + Batch Normalization+ Softmax	(128, K)

[a]K_{LSTM} is the output dimension of the LSTM from PoPPL. Interested readers can refer to [16] for details of PoPPL.
[b]C is the total number of weather-time conditions.

TABLE 2. Proposed statistical test of significance of weather-time conditions[a].

Class\Condition	1	2	…	C	Total
Class 1	l_{11}	l_{12}	…	l_{1C}	l_1
⋮	⋮	⋮	⋮	⋮	⋮
Class K	l_{K1}	l_{K2}	…	l_{KC}	l_K
Total	$\underline{n}_1$	$\underline{n}_2$	⋯	$\underline{n}_C$	$\underline{n}$

[a] $K \times C$ Contingency table of number of pedestrian arrivals from different destinations $\omega = 1, \ldots, K$ under different WT conditions $c = 1,2, \ldots, C$.

the GMU. The GMU is used to find an intermediate representation that fuses the two modalities, i.e. preliminary ID probabilities and WT embedding. First, the pedestrian intent probabilities and WT embedding are passed to individual tanh layers, each of which contains a neuron with hyperbolic tangent activation to encode the individual modalities. At the same time, a tied gate neuron learns the contribution of the two modalities, as shown in Fig. 2. The contributions $\mathfrak{z}_n$ and $(\mathbf{1} - \mathfrak{z}_n)$ obtained from the gate neuron will be multiplied in an elementwise manner to the output of the *tanh* layers of $\hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n)$ and $\boldsymbol{\Theta}\left(\boldsymbol{f}_{\boldsymbol{w},\boldsymbol{n}}, \boldsymbol{f}_{\boldsymbol{d},\boldsymbol{n}}\right)$, respectively. A special feature of this gate unit is that $\mathfrak{z}_n$ supports multivariate weighting. To use this feature, the output dimension of the two *tanh* layers can be modified to a common dimension matching each other. Finally, the fused multimodal representation will be passed to the final classifier for predicting the final class probability.

More precisely, the GMU can be described using the following set of equations:

$$\boldsymbol{h}_{\boldsymbol{n}}^{\boldsymbol{v}} = \tanh\left(\boldsymbol{W}_{\boldsymbol{v}} \cdot \hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n)\right), \tag{6a}$$

$$\boldsymbol{h}_{\boldsymbol{n}}^{\boldsymbol{e}} = \tanh\left(\boldsymbol{W}_{\boldsymbol{e}} \cdot \boldsymbol{\Theta}\left(\boldsymbol{f}_{\boldsymbol{w},\boldsymbol{n}}, \boldsymbol{f}_{\boldsymbol{d},\boldsymbol{n}}\right)\right), \tag{6b}$$

$$\mathfrak{z}_{\boldsymbol{n}} = \boldsymbol{\sigma}_{sgm}(\boldsymbol{W}_{\mathfrak{z}} \cdot \left[\hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n)^T, \boldsymbol{\Theta}\left(\boldsymbol{f}_{\boldsymbol{w},\boldsymbol{n}}, \boldsymbol{f}_{\boldsymbol{d},\boldsymbol{n}}\right)^T\right]^T), \tag{6c}$$

$$\boldsymbol{f}_{\boldsymbol{n}}^{\boldsymbol{fuse}} = \mathfrak{z}_{\boldsymbol{n}} \odot \boldsymbol{h}_{\boldsymbol{n}}^{\boldsymbol{v}} + (\mathbf{1} - \mathfrak{z}_{\boldsymbol{n}}) \odot \boldsymbol{h}_{\boldsymbol{n}}^{\boldsymbol{e}}, \tag{6d}$$

where $\boldsymbol{h}_{\boldsymbol{n}}^{\boldsymbol{v}}$ is the output of *tanh* layer for $\hat{\boldsymbol{p}}_{\boldsymbol{pre}}(\omega_n)$ for the n^{th} pedestrian. $\boldsymbol{h}_{\boldsymbol{n}}^{\boldsymbol{e}}$ is the output of *tanh* layer for Embedding. $\mathfrak{z}_{\boldsymbol{n}}$ is the output of the gate neuron. $\boldsymbol{f}_{\boldsymbol{n}}^{\boldsymbol{fuse}}$ denotes the fused representation. $\tanh(\boldsymbol{u}) = [\tanh(u_1), \tanh(u_2), \ldots]^T$ and $\tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$. $\boldsymbol{\sigma}_{sgm}(\boldsymbol{u}) = \left[\sigma_{sgm}(u_1), \sigma_{sgm}(u_2), \ldots\right]^T$ and $\sigma_{sgm}(u) = \frac{1}{1+e^{-u}}$. The Hadamard product operator is denoted as $\odot$. The set of unknown neural network weights to be learnt in the GMU are $\{\boldsymbol{W}_{\boldsymbol{v}}, \boldsymbol{W}_{\boldsymbol{e}}, \boldsymbol{W}_{\mathfrak{z}}\}$, which corresponds to the weights of tanh layer for the preliminary pedestrian intent probabilities, tanh layer for the WT embedding and Gate Neuron, respectively. A common dimension M is chosen for the two tanh layers in (6a) and (6b) so that they match the dimension of $\mathfrak{z}_{\boldsymbol{n}}$. Finally, a FC layer $\mathrm{FC}_{\mathrm{M,K}}()$ with input dimension M and output dimension K is used to learn the final ID class probabilities. $\hat{\boldsymbol{p}}_{\boldsymbol{F}}(\omega_n)$ is the predicted pedestrian intent probabilities obtained from the final classifier and it is given as

$$\hat{\boldsymbol{p}}_{\boldsymbol{F}}(\omega_n) = \boldsymbol{\sigma}_{\mathrm{Soft}}\left(\boldsymbol{\phi}_{\boldsymbol{BN}}\left(FC_{M,K}\left(\boldsymbol{f}_{\boldsymbol{n}}^{\boldsymbol{fuse}}\right)\right)\right), \tag{7}$$

where $\boldsymbol{\phi}_{\boldsymbol{BN}}$ and $\boldsymbol{\sigma}_{\mathrm{Soft}}$ are the batch normalization and softmax operations defined in (5c) and (5d), respectively. The preliminary and final classifiers will be jointly optimized as

$$\begin{aligned} L_T = {} & (1 - \lambda_P)\,\mathrm{L}_{\mathrm{focal}}\left(\omega, \hat{\boldsymbol{p}}_F(\omega)\right) \\ & + \lambda_P \mathrm{L}_{\mathrm{focal}}\left(\omega, \hat{\boldsymbol{p}}_{pre}(\omega)\right). \end{aligned} \tag{8}$$

where for simplicity, we drop the subscript n in (8). $\mathrm{L}_{\mathrm{focal}}\left(\omega, \hat{\boldsymbol{p}}_F(\omega)\right)$ and $\mathrm{L}_{\mathrm{focal}}\left(\omega, \hat{\boldsymbol{p}}_{pre}(\omega)\right)$ the losses for the final and preliminary classifiers, respectively. λ_P is a parameter controlling the ratio of the two losses. It is chosen as $\lambda_P = 0.5$ in this paper. To cater for possible class imbalance, the focal loss [28] is used

$$\begin{aligned} \mathrm{L}_{\mathrm{focal}}\left(\omega, \hat{\boldsymbol{p}}\right) = {} & -\frac{1}{NK}\sum_{n=1}^{N}\sum_{k=1}^{K} I_{k,n}\beta_k\left(1 - \hat{\boldsymbol{p}}_{k,n}\right)^{\gamma} \\ & \times \log\left(\hat{\boldsymbol{p}}_{k,n}\right), \end{aligned} \tag{9}$$

where $\hat{\boldsymbol{p}}_{k,n}$ is the predicted class probability for the $k - th$ class. $I_{k,n}$ is an indicator variable and $I_{k,n} = 1$ when the actual class is K. γ is a focusing factor and $\beta_k \in [0, 1]$ is a weighting factor. The final predicted class (i.e. intended destination) can be obtained as

$$\hat{\omega}_n = \max(\hat{\boldsymbol{p}}_F\left(\omega_{n,1}\right), \hat{\boldsymbol{p}}_F\left(\omega_{n,2}\right), \ldots, \hat{\boldsymbol{p}}_F(\omega_{n,K})), \tag{10}$$

where $\hat{\boldsymbol{p}}_F\left(\omega_{n,k}\right)$ is the softmax probability of the $k - th$ destination. $\hat{\boldsymbol{p}}_F(\omega) = [\hat{\boldsymbol{p}}_F\left(\omega_{n,1}\right), \hat{\boldsymbol{p}}_F\left(\omega_{n,2}\right), \ldots, \hat{\boldsymbol{p}}_F(\omega_{n,K})]^T$.

C. DESTINATION-ADAPTED TRAJECTORY PREDICTOR MODULE

After obtaining the final probabilities, the predicted trajectory can be obtained as

$$\hat{z}_n = \mathrm{DTP}_{k=\hat{\omega}_n}(s_n),$$

where $\hat{z}_n$ is the predicted trajectory for the chosen class. $\mathrm{DTP}_{k=\hat{\omega}_n}(s_n)$ is the chosen destination trajectory baseline model based on predicted class $\hat{\omega}_n$. The baseline model

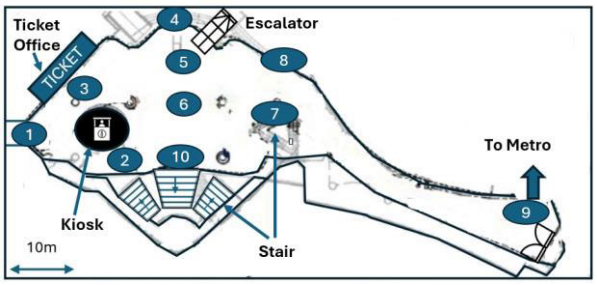


FIGURE 3. Initialization centroids for k-means clustering, $K = 10$ classes added to 1/F floor plan of Osaka ATC Centre [29]. Important functional objects (i.e. ticket office, escalator, kiosk and stairs) are redrawn. The historical weather on 22nd May 2013 (sunny) and 29th September (cloudy), 2013 was obtained from [45].

TABLE 3. List of initialization centroids for Osaka ATC Centre 1/F.

	$K = 10$	[1]$K = 9$
Exit to ferry pier	Centroid 1	Centroid 1
Exit to ticket office	Centroid 3	Centroid 3
Information Kiosk	Centroid 2	Centroid 2
Central Square	Centroid 6	Centroid 6
Stairs to G/F	Centroid 10	
Stairs to 2/F and floor guide	Centroid 7	Centroid 7
Stairs to the mall	Centroid 8	Centroid 8
Exit to metro station	Centroid 9	Centroid 9
Stairs to the mall	Centroid 4	Centroid 4
Escalator	Centroid 5	Centroid 5

[1]According to the statistical test in Table IV, classes 6 and 10 are merged.

is chosen as sub-LSTM (PoPPL-def) in PoPPL [16] for the sake of comparison. The sub-LSTM (PoPPL-def) is an encoder-decoder LSTM with two hidden layers. In the next sub-section, the proposed statistical test will be presented.

D. STATISTICAL TEST FOR WEATHER-TIME CONDITIONS

The proposed statistical test can be used to establish the minimum required samples for each cluster and to quantify whether it is necessary to treat the pedestrian movement pattern in different periods and weathers as different groups and use different trajectory models to describe their behavior. More precisely, Table 2 shows a $K \times C$ contingency table summarizing the number of pedestrians arriving to K destinations under C different weather-time (WT) conditions. The following null hypothesis is proposed:

$$H_0 : \text{The WT condition does not affect the choice of destination.} \tag{11}$$

If the null hypothesis is true, then the observed number of pedestrians should not deviate significantly from the expected counts across different WT conditions. According to the χ^2 test, the minimum number of expected samples/trajectories required for each cluster k under condition c is

$$e_{kc} = \frac{l_K \times \underline{n}_C}{\underline{n}} \geq 5, \tag{12}$$

TABLE 4. Number of pedestrian arrivals during peak hour (12:00-16:59). off-peak (09:00 –11:59, 17:00 – 20:00), sunny and cloudy for Osaka ATC dataset (K = 10).

	Weather-Time conditions				Total
	A	B	C	D	
Class 1	645	601	1135	1722	4103
Class 2	71	102	113	256	542
Class 3	25	46	123	230	424
Class 4	625	953	2010	3912	7500
Class 5	75	106	281	445	907
Class 7	126	186	439	667	1418
Class 8	653	1044	1226	2303	5226
Class 9	938	1072	2637	3436	8083
Class 6	1	2	6	21	**30**
Class 10	20	38	55	190	303
Total	**3179**	4150	8025	13182	28536
Min. sample before merging (K=10)	Applying Eqn. (12) $e_{61} = \frac{l_6 \underline{n}_1}{\underline{n}} = \frac{(30)(3179)}{28536} = 3.34 < 5$ **(not satisfied)**				
Min. sample after merging (K=9)	Applying Eqn. (12) $e_{61} = \frac{l_6 \underline{n}_1}{\underline{n}} = \frac{(\mathbf{30+303})(3179)}{28536} = 37.09 > 5$ **(satisfied)**				
χ^2 after merging ($K = 9$)	588.64 (degree of freedom 24)				
log(p-value) after merging ($K = 9$)	-104.8395 < log (0.05) (significant)				

A: cloudy+off-peak B: cloudy+peak C: sunny+off-peak D: sunny+peak
[a]Peak hour: 12:00-16:59, Off peak: 09:00– 11:59, 17:00 – 20:00
[b]Peak hour is chosen from 12pm to 5pm because the total number of pedestrians during that period surpasses the remaining hours of the day.

where $l_k = \sum_{c=1}^{C} l_{kc}$, $\underline{n}_c = \sum_{k=1}^{K} l_{kc}$ and $\underline{n} = \Sigma_{k=1}^{K} \Sigma_{c=1}^{C} l_{kc}$. l_{kc} is the number of observed trajectories/samples in the $k-th$ destination and $c-th$ condition.

Once the clusters are established, the test statistic for WT condition reads

$$\chi^2_{obs} = \sum_{c=1}^{C} \sum_{k=1}^{K} \frac{(o_{kc} - e_{kc})^2}{e_{kc}}, \tag{13}$$

where o_{kc} is the actual observed number of pedestrians in condition c and destination k. The p-value of the test is given as

$$p = Pr(\chi^2 \geq \chi^2_{obs,j} | H_0), \tag{14}$$

where the test statistic follows a χ^2 distribution with $(C-1)(K-1)$ degree of freedom. At a significance level of 0.05 [31], the null hypothesis will be rejected when the p-value is smaller than 0.05 and it will suggest the difference between the proportion of pedestrian arrival under different conditions and origins is statistically significant.

IV. RESULTS AND ANALYSIS

For illustrative purposes, the Osaka Asia and Pacific Trade Center (ATC) dataset (Dražen et al. 2013) is considered. The Osaka ATC is a transportation hub linking the Sunflower inter-city Ferry pier to the Osaka City Metro. It contains a multi-entertainment complex and a conference center. The trajectories were collected at 1/F of ATC using 3D range

sensors. The full dimension is over 140 $m \times 60$ m. Trajectories from 0900 to 2000 on 22nd May, 2013 (sunny) and 29th September, 2013 (cloudy) are chosen. Trajectories that are too short are removed (i.e. same cluster for origin and destination) as they may be a result of occlusion or tracking loss of the 3D range sensor. After resampling, the trajectory length $L+L' = 40$. The total number of pedestrian trajectories after pre-processing are 7329 on 22nd May, 2013 (sunny) and 21207 on 29th September, 2013 (cloudy) and respectively. Hence, the total number of pedestrian/ trajectories are 28536. Each pedestrian contains only 1 trajectory.

A. CHOICE OF CLUSTER

A general rule of thumb to choose the number of classes is to study the number of possible entrances/exits of the floor plan [16], [17]. Fig. 3 shows the floor plan of the Osaka ATC Center (1/F). Following this notion, key entrances and exits are chosen as the initialization centroids as in Fig. 3. Table 3 shows the list of initialization centroids.

B. STATISTICAL ANALYSIS OF TIME-OF-DAY AND WEATHER CONDITIONS

In this sub-section, we shall test the significance of time-of-day and weather conditions using the proposed statistical test.

Table 4 shows the number of observed pedestrian arrival during peak hour (12:00-16:59), off-peak, sunny and rainy conditions for $K = 10$. Using (12), it was found that class ω_6 does not meet the minimum sample requirement. Hence, using the centroid linkage criterion, ω_6 is merged with ω_{10}. The observed χ^2_{obs} computed using (13) is 588.64 (degree of freedom 24) and the log(p-value) is -104.8395, which is statistically significant under the typical significance level of 0.05, Ross [31]. This suggests there is a significant deviation in the pedestrian counts across the different clusters under the different conditions. Hence, the proposed approach should be used to model the pedestrian trajectory patterns under the different conditions. Next, we shall evaluate the performance of the various algorithms.

C. BASELINE AND METRIC

To evaluate the performance of the proposed approach, we compare the proposed WTTFNet with the following algorithms:

1. Linear Model: A simple linear model with a hidden layer (nn.linear in Pytorch) [46] is used to predict the trajectories.
2. Vanilla LSTM: The sub-LSTM in PoPPL-def is used. It employs an encoder-decoder LSTM with 2 hidden layers fitting all the trajectories. The implementation follows the Github codes [16].
3. PoPPL [16]: The sub-LSTM model is employed together with route class clustering. The implementation follows the Github codes. Following the previous statistical analysis, $K = 9$ destinations were chosen. Route class clustering divides all trajectories according to all combinations of all 9 origins and 9 destinations for training trajectory models.
4. Proposed WTTFNet: For fair comparison, we adopt the same baseline model as in PoPPL, as shown in Fig. 1. However, the proposed destination-driven clustering and proposed WTTFNet are used. Hyperparameters same as the authors are adopted for the PoPPL baseline model. For the number of destinations, $K = 9$ is chosen as in the previous analysis.

For evaluating the quality of trajectory prediction, the average displacement error (ADE) is the average Euclidean distance between all the actual and all predicted coordinates over all trajectories. The FDE is the average Euclidean distance between the final destination of the predicted and actual trajectories. They are given as

$$\text{ADE} = \frac{1}{N_T L'} \sum_{n=1}^{N_T} \sum_{t=1}^{L'} \left\| \begin{pmatrix} x_{n,L+t} \\ y_{n,L+t} \end{pmatrix} - \begin{pmatrix} \hat{x}_{n,L+t} \\ \hat{y}_{n,L+t} \end{pmatrix} \right\|_2, \tag{15a}$$

$$\text{FDE} = \frac{1}{N_T} \sum_{n=1}^{N_T} \left\| \begin{pmatrix} x_{n,L+L'} \\ y_{n,L+L'} \end{pmatrix} - \begin{pmatrix} \hat{x}_{n,L+L'} \\ \hat{y}_{n,L+L'} \end{pmatrix} \right\|_2, \tag{15b}$$

where $||.||_2$ denotes the Euclidean distance. N_T is the total number of testing samples. $(x_{n,t}, y_{n,t})$ is the actual coordinate and $(\hat{x}_{n,t}, \hat{y}_{n,t})$ is the predicted coordinate of the $n-th$ pedestrian's trajectory. The accuracy of the destination classification is evaluated using classification accuracy (ACC) and Cohen's Kappa (κ). They are given as

$$\text{ACC} = \frac{1}{N_{Test}} \sum_{k=1}^{K} CM[i,i], \tag{16a}$$

$$\kappa = \frac{N_T \sum_{i=1}^{K} CM\,[i,i] - \sum_{i=1}^{K} C_T\,[i]\, C_P[i]}{N_T^2 - \sum_{i=1}^{K} C_T\,[i]\, C_P[i]}, \tag{16b}$$

where $CM[i,j] = \sum_{n=1}^{N_{Test}} I(\omega_n = i \,\&\, \hat{\omega}_n = j)$ is the total number of counts of having the actual class i and predicted class j. I is the indicator function. $C_T\,[i] = \sum_{j=1}^{K} CM[i,j]$ and $C_P\,[j] = \sum_{i=1}^{K} CM[i,j]$. While classification accuracy is commonly used to describe the generic performance, Cohen's Kappa is used more frequently for scenarios with possible class imbalance. The following relative metrics, rd are used to compare between different algorithms,

$$rd = \frac{(d - d_{REF})\,(-1)^m}{d_{REF} + \epsilon} \times 100\%, \tag{17}$$

where d can be any metrics, such as the ADE, FDE, ACC and κ. d_{REF} is the performance of the reference method. $\epsilon = 10^{-8}$ is a small constant added to denominator to avoid division by zero. m is a parameter defining metric type. $m = 0$ is used for maximizing metrics with larger value indicating better performance, whereas $m = 1$ is used for loss metrics with smaller value indicating better performance.

D. EXPERIMENTAL SETUP

The Google Colab Tesla T4 Graphics Processing Unit (GPU) notebook with 16GB GPU memory and 17 GB of system memory is used for evaluation. In the experiment, each observed trajectory has a duration of 20 time-instants and an algorithm will predict the trajectory for the next 20 time-instants. Fig. 3 shows the validation protocol following the

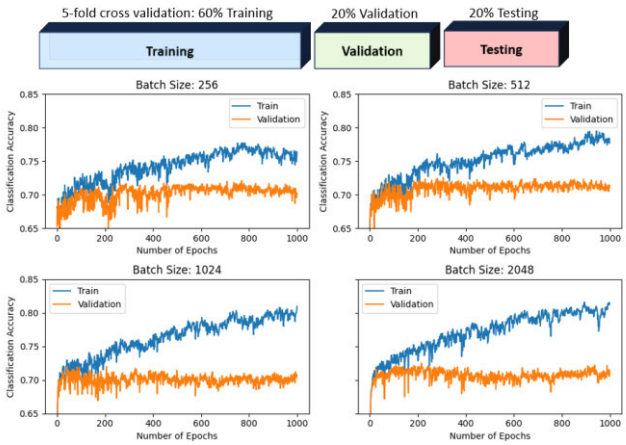


FIGURE 4. Validation protocol and learning curves for various batch sizes. Stratified 5-fold cross validation (CV) is used.

validation strategy in [16]. Stratified 5-fold cross validation (CV) is employed. Due to stratification and possible chances that the total number of samples is indivisible by 5, the number of samples across folds may vary slightly. Three-folds (∼60%), one-fold (∼20%), and one-fold (∼20%) are used for training, validation, and testing, respectively.

1) BATCH SIZE AND STOPPING CRITERION

Fig. 4 shows the training and validation curves for the proposed WTTFNet under batch sizes 256, 512,1024 and 2048. For batch size 256, the validation curve is quite noisy and fluctuates rapidly and hence it is not considered. For batch sizes 512, 1024 and 2048, the training accuracy starts to level off around epoch 100 but the validation accuracy remains roughly around a certain range. This suggests more epochs do not necessarily lead to better validation performance. Hence, 1000 epochs are chosen as stopping criterion. Overall, batch size 1024 attained the lowest variance in validation accuracy and hence it is chosen. For each CV fold, the model obtained at the epoch attaining the best validation accuracy is chosen and is used to evaluate the testing data.

2) HYPERPARAMETERS

Hyperparameters same as the PoPPL are adopted for the baseline classifier and trajectory models. Dropout parameter of 0.5 and hidden size of 128 are adopted. For the proposed WTTFNet, the weighing factor in the focal loss is chosen as $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_K]^T$, $\beta_k = (\frac{N/N_k}{\sum_{k=1}^{K} N/N_k})$, where N is the total number of training samples, N_k is the number of training samples of class k, K is the total number of classes. The focusing parameter is chosen as $\gamma = 2$. The ratio between the preliminary and final loss in (8) is chosen as $\lambda_P = 0.5$.

E. EXPERIMENTAL RESULTS

In this sub-section, the proposed WTTFNet is compared against various algorithms. Since the proposed WTTFNet can be attached to arbitrary deep neural network baseline models, the PoPPL is adopted as baseline for illustration. In general, other deep neural network based intent classifier, such as transformers, can be adopted as the baseline model. Since the PoPPL is a technique that combined clustering and LSTM, we also compared with Vanilla LSTM.

TABLE 5. Trajectory prediction performance of various algorithms.

Metric \ Model	Linear Model	Vanilla LSTM	PoPPL[c] (Original)	Proposed WTTFNet[c] (A)	Proposed WTTFNet[c] (B)
ACC (%)	N/A	N/A	58.18%	71.50%	**71.95%**
ADE(m)	13.28	6.263	6.488	5.93	**5.894**
FDE(m)	22.84	10.687	11.266	10.42	**10.315**
rACC (%)[a,d]	N/A[b]	N/A[b]	Reference	22.89%	**23.67%**
rADE (%)[a,e]	-104.69%	3.47%		8.58%	**9.16%**
rFDE (%)[a,e]	-103.46%	3.72%		6.13%	**7.07%**

A: Proposed WTTFNet without weather-time (WT) information
B: Proposed WTTFNet with WT information
[a] The relative improvement metrics are defined as in (17).
[b]The Linear model and Vanilla LSTM does not contain a destination/route classifier. Hence, classification performance is not applicable.
[c]Both the PoPPL and the proposed WTTFNet select the specialized trajectory models using a classifier.
[d]rACC measures relative increase in ACC.
[e]rADE and rFDE measures relative reduction in ADE and FDE, respectively.

TABLE 6. Trajectory prediction performance of various algorithms.

Metric \ Model	PoPPL[c] (Original)	PoPPL[d]+ Focal Loss	Proposed WTTFNet[e] Without WT info	Proposed WTTFNet[e] With WT info
ACC[a]	58.18%	68.84%	71.50%	**71.95%**
κ[a]	51.73%	62.49%	65.89%	**66.28%**
rACC[b]	Reference	18.32%	22.89%	**23.67%**
rκ[b]		20.80%	27.37%	**28.13%**

[a] ACC: Accuracy; κ: Cohen's Kappa
[b] Relative improvements of ACC and κ.
[c] PoPPL (Original): Original PoPPL optimized with loss function using cross entropy.
[d] PoPPL (FL): PoPPL optimized using focal loss (FL)
[e] Proposed WTTFNet: It is made up of PoPPL + FL + Deep supervision + WT information incorporated using GMU
[f]Since performance may vary across different baseline models, the PoPPL is adopted as the baseline model of the proposed approach for fair comparison. In general, other baseline models can also be applied.

Table 5 shows the overall performance of all algorithms. The proposed WTTFNet performed better than the original PoPPL, Vanilla LSTM and the linear model for all cases considered. Particularly, the proposed WTTFNet surpasses the original PoPPL 23.67% in classification accuracy, 9.16% reduction in ADE and 7.07% reduction in FDE. Significant p-values of ($p < 10^{-16}$) are attained for improvement in classification accuracy (McNemar's test [30]), ADE and FDE (one-sided Mann–Whitney U tests [32]).

1) ABLATION TEST

The intended destination classifier of the proposed WTTFNet is made up of i) baseline model ii) focal loss iii) deep supervision (preliminary and final classifiers co-trained with

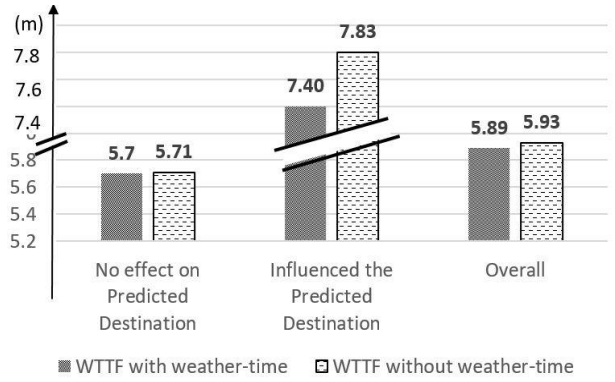


FIGURE 5. Average Displacement Error (ADE) of the proposed approach with/without the incorporation of weather-time information. Significant reduction in ADE can be observed (7.8m to 7.4m) for the significant 3008 pedestrians out of all 28536 pedestrians.

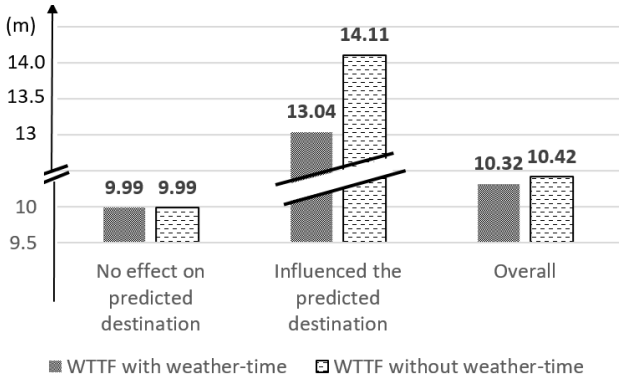


FIGURE 6. Final Displacement Error (FDE) of the proposed approach with/without the incorporation of weather-time information. Significant reduction in FDE can be observed (14.11m to 13.04m) for the significant 3008 pedestrians out of all 28536 pedestrians.

joint loss function) and iv) incorporation of WT information using GMU. To study the incremental contribution of each component of the proposed novel WTTFNet and show the role of weather and time-of-day in improving the prediction, we consider Table 6, which compare the following four different settings:

1. PoPPL (baseline model): The original PoPPL with entropy loss. In general, other baseline models can be used.
2. PoPPL (baseline model) + FL: PoPPL modified with Focal Loss.
3. WTTFNet without WT information (second last column of Tables 5 and 6): GMU is bypassed and WT information is not incorporated. Deep supervision is used to co-train the preliminary and final classifiers.
4. WTTFNet with WT information (final column of Tables 5 and 6): Between the preliminary and final classifiers, the GMU is inserted and the WT information is fused with the preliminary pedestrian intent probabilities at the GMU.

Comparing between PoPPL and PoPPL+FL (Setting 1 vs 2), it can be seen that the use of focal loss improves the ACC as it helps to tackle the class imbalance existed among the clusters. After adding the proposed WTTFNet (Setting 2 vs Setting 3), even without the WT information, around 4% relative improvement in ACC is observed. This suggests even when the GMU is bypassed and WT information is not supplied, deep supervision employed in the WTTFNet is useful in refining both the preliminary and final classifiers optimized using auxiliary and final loss functions based on focal loss. This leads to improved classification accuracy (Table 6), reduction in ADE and FDE (Table 5).

Finally, to study the role of weather and time-of-day in improving the performance, we compare WTTFNet without/with WT information (Setting 3 vs Setting 4). We can observe that the best performance (highest classification accuracy, lowest ADE and FDE) can be attained in Tables 6 and 5, respectively, after incorporation of WT information into the proposed WTTFNet, which suggests the usefulness in adding WT information in prediction.

Overall, the ACC increased from 71.5% to 71.95% after adding WT information in the proposed WTTFNet. To validate its statistical significance, we performed a McNemar's test and a significant p-value ($p = 0.0196 < 0.05$) was attained. This suggests the improvement from 71.5% to 71.95% in ACC is very unlikely to be solely due to random under the large sample size of 28536 pedestrians. What follows, the McNemar's test also identifies 3008 pedestrians out of 28536 to have deviation in identified destination classes after WT information is added, this prompted us to further analyze those two different groups of pedestrians in next subsection.

2) QUANTITATIVE ANALYSIS OF THE ROLE OF WEATHER AND TIME-OF-DAY

In this section, further analysis on the role of weather and time-of-day in improving the prediction performance is studied. Following the significance p-value obtained for the McNemar's test in previous sub-section, which suggests that there is significant improvement in classification accuracy after adding WT information to the proposed approach. Moreover, 3008 pedestrians were found to have significant improvement after WT information were added. This motivates us to analyze the average displacement error (ADE) and final displacement error (FDE) of the 3008 pedestrians.

Figs. 5 and 6 compare the ADE and FDE of the proposed approach under two settings, respectively: with/without the incorporation of WT information. "No effect on Destination" means the predicted destination are same under the both settings, whereas "Influenced the Predicted Destination" means the predicted destination was altered after incorporating the WT information.

Fig. 5 shows the ADE of the proposed approach with/without WT information incorporated. From the figure, it can be shown that similar ADE was attained when the WT information has no effect on the predicted destination. On the other hand, if the predicted destination changed because of

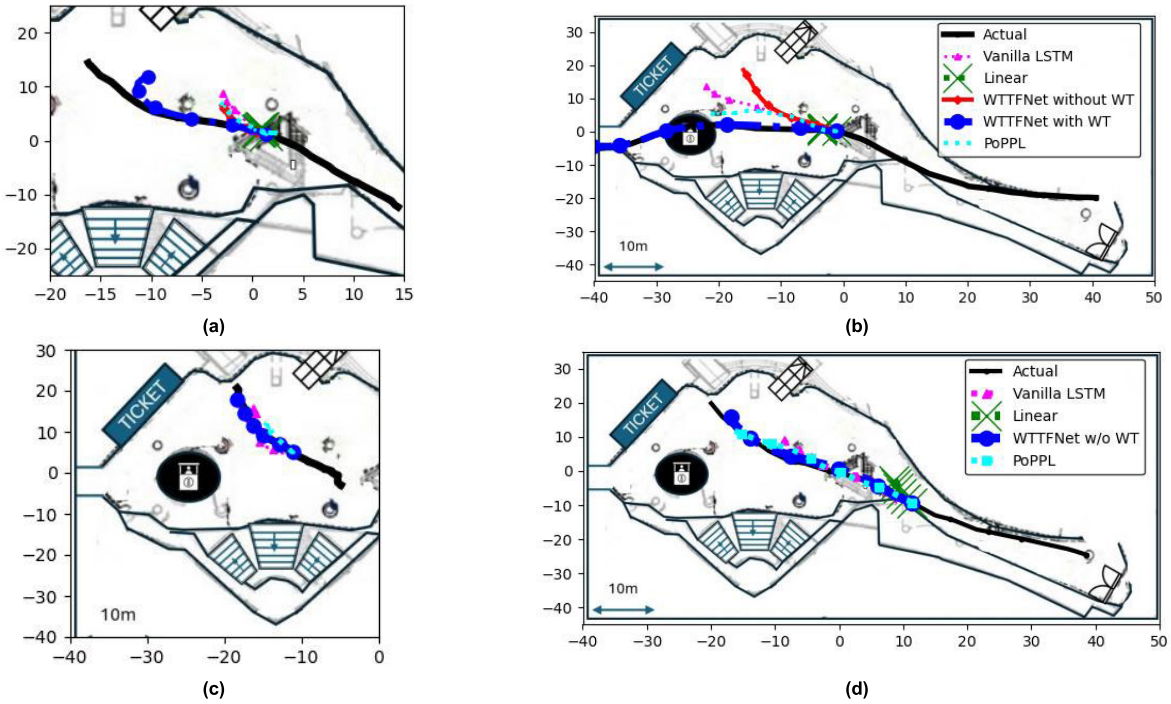


FIGURE 7. Illustration of predicted trajectories, where the weather-time condition has (a,b) significant influence on destination (chosen from the 3008 significant pedestrians), and (c,d) no influence on destination (chosen from remaining pedestrians). The first half of the trajectory (denoted in black) is used to predict the latter half of the trajectory. Since the two lines of WTTFNet with/ without WT overlapped in (c) and (d), both settings are merged to one line.

the varying weather (Influenced the predicted destination), the proposed WTTF with WT information incorporated will attain lower ADE (7.4083m) in compared to without WT information (7.83m). One-sided Mann–Whitney U test was used to test the significance in ADE reduction (7.83m to 7.4083m after adding WT information) and a p-value of $p = 0.0203 < 0.05$ was attained, suggesting the significance in performance improvement for these pedestrians considered.

Fig. 6 shows the FDE under the two settings (with/without WT information) were compared for the proposed approach. Similar to the observation in the previous comparison, same FDE was attained when the WT information has no effect on the predicted destination (FDE = 9.99m) and improved FDE (reduction from 14.11m to 13.04m) for the proposed WTTF approach when it changes the predicted destination after incorporating WT information. One-sided Mann–Whitney U test was used to test the significance in FDE reduction (14.11m to 13.04m after adding WT information) and a p-value of $p = 0.00533 < 0.05$ was attained, suggesting the significance in performance improvement for these pedestrians considered.

Overall, 5.47% (7.8m to 7.4m) and 7.58% (14.11m to 13.04m) improvement in ADE and FDE reduction were obtained for the 3008 pedestrians, and the reduction is found significant according to one-sided Mann–Whitney U tests. ($p = 0.0203$ (<0.05) and $p = 0.00533$ (<0.05) for ADE and FDE, respectively). For the remaining pedestrians, similar ADE and FDE performance was observed for pedestrians with no effect, because they have the same predicted destination under two settings (with/without WT information).

3) QUALITATIVE ANALYSIS OF THE ROLE OF WEATHER AND TIME-OF-DAY

To illustrate the usefulness of adding WT information in the proposed WTTFNet, we consider four different cases, where Figs. 7(a) and (b) are extracted from the significant 3008 pedestrians and Figs 7(c) and (d) are extracted from the remaining pedestrians, whose destination was not affected by weather-time conditions.

Comparing between the proposed WTTFNet and other algorithms, the proposed WTTFNet (solid blue line with dots) generally aligns the best with the actual trajectory (solid black). In particular, the linear model, vanilla LSTM and PoPPL diverged inferiorly in Figs. 7(a) and 7(b).

To study the role of weather and time-of-day, we compare between the two different settings of the proposed WTTFNet: with/without WT information. From Figs 7(a) and 7(b), the WTTFnet with WT information (solid blue line with dots) aligns much better than the counterpart without WT information (solid red line with diamonds), which diverges in the middle of the path. For the remaining non-significant pedestrians, both settings nearly the same performance in Figs. 7(c) and 7(d) and hence only one of them are plot on the graphs.

Overall, the quantitative (Figs. 5 and 6) and qualitative (Fig. 7) analyses show that weather-time information

helps to improve prediction performance significantly for the 3008 cases considered. The proportion of 3008 out of 28536 was also statistically significant according to the McNemar's test, suggesting that these 3008 pedestrians showing improved performance out of 28536 cases were very unlikely a random event. This suggests the proposed approach may serve as an attractive approach for incorporating WT information to improve pedestrian trajectory prediction and it also serves as a systematic approach to test the significance of WT conditions.

V. CONCLUSION

A new deep WTTFNet has been presented. Experimental results using the Osaka ATC dataset [3] show that the proposed approach attained better performance than other state-of-the-art methods considered under varying weather-time conditions. A statistical test is also used to establish the significance of time-of-day and weather conditions. The proposed refinement framework can be adopted on other baseline models to improve these performance under varying weather-time conditions.

REFERENCES

[1] J. Kantorovitch, J. Väre, V. Pehkonen, A. Laikari, and H. Seppälä, "An assistive household robot–doing more than just cleaning," *J. Assistive Technol.*, vol. 8, no. 2, pp. 64–76, 2014.

[2] S. Song, J. Baba, J. Nakanishi, Y. Yoshikawa, and H. Ishiguro, "Teleoperated robot sells toothbrush in a shopping mall: A field study," in *Proc. Extended Abstracts CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–6.

[3] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 51, no. 5, p. 4282, 1995.

[4] M. Moussaïd, D. Helbing, and G. Theraulaz, "How simple rules determine pedestrian behavior and crowd disasters," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 17, pp. 6688–6884, 2011.

[5] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

[6] Y. Yao, E. Atkins, M. Johnson-Roberson, R. Vasudevan, and X. Du, "BiTraP: Bi-directional pedestrian trajectory prediction with multi-modal goal estimation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 1463–1470, Apr. 2021.

[7] B. Liu, E. Adeli, Z. Cao, K.-H. Lee, A. Shenoi, A. Gaidon, and J. C. Niebles, "Spatiotemporal relationship reasoning for pedestrian intent prediction," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3485–3492, Apr. 2020.

[8] J. Wang, H. Sang, W. Chen, and Z. Zhao, "VOSTN: Variational one-shot transformer network for pedestrian trajectory prediction," *Phys. Scripta*, vol. 99, no. 2, Feb. 2024, Art. no. 026002.

[9] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2255–2264.

[10] Y. Yuan, X. Weng, Y. Ou, and K. Kitani, "AgentFormer: Agent-aware transformers for socio-temporal multi-agent forecasting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9813–9823.

[11] Z. Lv, X. Huang, and W. Cao, "An improved GAN with transformers for pedestrian trajectory prediction models," *Int. J. Intell. Syst.*, vol. 37, no. 8, pp. 4417–4436, Aug. 2022.

[12] Q. Du, X. Wang, S. Yin, L. Li, and H. Ning, "Social force embedded mixed graph convolutional network for multi-class trajectory prediction," *IEEE Trans. Intell. Vehicles*, 2024, doi: 10.1109/TIV.2024.3352180.

[13] Z. He, T. Zhang, W. Wang, and J. Li, "A deep pedestrian trajectory generator for complex indoor environments," *Trans. GIS*, vol. 28, no. 2, pp. 411–432, Apr. 2024.

[14] Y. Han, C. S. Tucker, T. W. Simpson, and E. Davidson, "A data mining trajectory clustering methodology for modeling indoor design space utilization," in *Proc. Int. Design Eng. Tech. Conferences Comput. Inf. Eng. Conf.*, vol. 55898, 2013, Art. no. V03BT03A017, doi: 10.1115/DETC2013-12690.

[15] M. D'Orazio, G. Bernardini, and E. Quagliarini, "How to restart? An agent-based simulation model towards the definition of strategies for COVID-19 'second phase' in public buildings," 2020, *arXiv:2004.12927*.

[16] H. Xue, D. Q. Huynh, and M. Reynolds, "PoPPL: Pedestrian trajectory prediction by LSTM with automatic route class clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 77–90, Jan. 2021.

[17] A. K. F. Lui, Y. H. Chan, and K. Hung, "Functional objects in urban walking environments and pedestrian trajectory modelling," *Sensors*, vol. 23, no. 10, p. 4882, May 2023.

[18] A. T. Steele, "Weather's effect on the sales of a department store," *J. Marketing*, vol. 15, no. 4, pp. 436–443, 1951.

[19] N. Rose and L. Dolega, "It's the weather: Quantifying the impact of weather on retail sales," *Appl. Spatial Anal. Policy*, vol. 15, no. 1, pp. 189–214, 2022.

[20] E. Chung, O. Ohtani, H. Warita, M. Kuwahara, and H. Morita, "Effect of rain on travel demand and traffic accidents," *Proc. IEEE Intell. Transp. Syst.*, Sep. 2005, pp. 13–16.

[21] H. Ren, Y. Song, S. Li, and Z. Dong, "Two-step optimization of urban rail transit marshalling and real-time station control at a comprehensive transportation hub," *Urban Rail Transit*, vol. 7, pp. 257–268, Sep. 2021.

[22] Y. Shi, J. Xu, H. Zhang, L. Jia, and Y. Qin, "Walking model on passenger in merging passage of subway station considering overtaking behavior," *Phys. A, Stat. Mech. Appl.*, vol. 585, Jan. 2022, Art. no. 126436.

[23] N. Tsiamitros, T. Mahapatra, I. Passalidis, K. Kailashnath, and G. Pipelidis, "Pedestrian flow identification and occupancy prediction for indoor areas," *Sensors*, vol. 23, no. 9, p. 4301, Apr. 2023.

[24] J. Peng, F.-L. Peng, N. Yabuki, and T. Fukuda, "Factors in the development of urban underground space surrounding metro stations: A case study of osaka, Japan," *Tunnelling Underground Space Technol.*, vol. 91, Sep. 2019, Art. no. 103009.

[25] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, "Are they going to cross? A benchmark dataset and baseline for pedestrian crosswalk behavior," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Venice, Italy, Oct. 2017, pp. 206–213.

[26] J. Arevalo, T. Solorio, M. Montes-y-Gómez, and F. A. González, "Gated multimodal units for information fusion," 2017, *arXiv:1702.01992*.

[27] C. Li, M. Z. Zia, Q.-H. Tran, X. Yu, G. D. Hager, and M. Chandraker, "Deep supervision with intermediate concepts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1828–1843, Aug. 2019.

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[29] D. Brščić, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-D range sensors," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 6, pp. 522–534, Nov. 2013.

[30] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, Jun. 1947.

[31] S. M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*. New York, NY, USA: Academic, 2020.

[32] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Statist.*, vol. 18, no. 1, pp. 50–60, Mar. 1947.

[33] V. Mehta and J. K. Bosson, "Revisiting lively streets: Social interactions in public space," *J. Planning Educ. Res.*, vol. 41, no. 2, pp. 160–172, 2021.

[34] X. Yang, R. Zhang, F. Pan, Y. Yang, Y. Li, and X. Yang, "Stochastic user equilibrium path planning for crowd evacuation at subway station based on social force model," *Phys. A, Stat. Mech. Appl.*, vol. 594, May 2022, Art. no. 127033.

[35] B. I. Sighencea, R. I. Stanciu, and C. D. Căleanu, "A review of deep learning-based methods for pedestrian trajectory prediction," *Sensors*, vol. 21, no. 22, p. 7543, Nov. 2021.

[36] C. Zhang and C. Berger, "Pedestrian behavior prediction using deep learning methods for urban scenarios: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 10, pp. 10279–10301, Oct. 2023.

[37] D. Chowdhury, L. Santen, and A. Schadschneider, "Statistical physics of vehicular traffic and some related systems," *Phys. Rep.*, vol. 329, nos. 4–6, pp. 199–329, 2000.

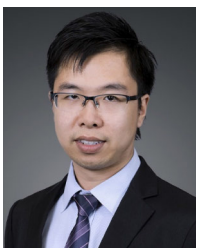
[38] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 961–971.
[39] A. Rasouli, I. Kotseruba, T. Kunic, and J. Tsotsos, "PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 6261–6270.
[40] F. Piccoli, R. Balakrishnan, M. J. Perez, M. Sachdeo, C. Nuñez, M. Tang, K. Andreasson, K. Bjurek, R. Dass Raj, E. Davidsson, C. Eriksson, V. Hagman, J. Sjöberg, Y. Li, L. Srikar Muppirisetty, and S. Roychowdhury, "FuSSI-net: Fusion of spatio-temporal skeletons for intention prediction network," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2020, pp. 68–72.
[41] Y. Li, C. Zhang, J. Zhou, and S. Zhou, "POI-GAN: A pedestrian trajectory prediction method for service scenarios," *IEEE Access*, vol. 12, pp. 53293–53305, 2024, doi: 10.1109/ACCESS.2024.3387698.
[42] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1349–1358.
[43] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14424–14432.
[44] B. Yang, W. Zhan, P. Wang, C. Chan, Y. Cai, and N. Wang, "Crossing or not? Context-based recognition of pedestrian crossing intention in the urban environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5338–5349, Jun. 2022.
[45] Time and Date AS. (2024). *Historical Weather Data from Osaka International Airport, Japan*. Accessed: Mar. 2, 2024. [Online]. Available: https://www.timeanddate.com
[46] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.

HO CHUN WU (Member, IEEE) received the B.Eng. degree in electrical engineering from the University of New South Wales, Sydney, NSW, Australia, in 2008, and the M.Sc. degree in electrical and electronic engineering and the Ph.D. degree from The University of Hong Kong, Hong Kong, in 2009 and 2013, respectively. He is currently a Lecturer with the School of Science of Technology, Hong Kong Metropolitan University, Hong Kong. His research interests include big data analytics, pattern recognition, bioinformatics, biomedical signal processing, time series analysis, and smart grids.

ESTHER HOI SHAN LAU received the B.Sc. degree in computer science from the City University of Hong Kong, in 2023, where she is currently pursuing the M.Sc. degree in biostatistics. She is also a Research Assistant with the School of Nursing and Health Studies, Hong Kong Metropolitan University.

PAUL CHUN HO YUEN received the B.Sc. degree in data science from the School of Science and Technology, Hong Kong Metropolitan University, in 2022. He is currently pursuing the M.Sc. degree in artificial intelligence and big data computing with The Hong Kong Polytechnic University. From 2022 to 2023, he was an AI Developer at ATTA Technologies Ltd. He is also a Research Assistant at Hong Kong Metropolitan University.

KEVIN HUNG (Senior Member, IEEE) received the B.Sc. degree from Queen's University, Canada, and the M.Phil. and Ph.D. degrees from The Chinese University of Hong Kong (CUHK). He is currently an Associate Professor and the Head of the Department of Electronic Engineering and Computer Science, School of Science and Technology, Hong Kong Metropolitan University (HKMU). Prior to joining HKMU, he was the Assistant Project Manager at the Joint Research Centre for Biomedical Engineering, CUHK, and an Engineer at Medical Device Company. He is the principal investigator of several projects funded by both the government and the university. His research interests include mobile health, wearable sensors, bio-signal processing, bio-system simulation, bio-medical informatics, and engineering education. In addition to his academic roles, he is a founding Officer of the IEEE Engineering in Medicine and Biology Society (EMBS) Hong Kong–Macau Joint Chapter, and served as its Chair, in 2010. He is also the Vice Chair of the IEEE Hong Kong Section, a founding Counsellor of the IEEE HKMU Student Branch, an Immediate Past Chair of the Electronics and Communications Section at IET Hong Kong, a Committee Member of the IET Hong Kong Branch, and the Honorary Secretary of the Chinese Institute of Electronics, Hong Kong.



JOHN KWOK TAI CHUI (Member, IEEE) received the B.Eng. degree in electronic and communication engineering (business intelligence minor) and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong. He is currently with the Department of Electronic Engineering and Computer Science, School of Science and Technology, Hong Kong Metropolitan University, as an Assistant Professor. He had industry experience as a Senior Data Scientist with the Internet of Things (IoT) Company. He has published more than 100 research works, including edited books, book chapters, journal papers, and conference papers. His research interests include computational intelligence, cyber security, biomedical signal processing, machine learning, and optimization. He was a recipient of the 2nd Prize Award (Postgraduate Category) of the 2014 IEEE Region 10 Student Paper Contest and the Best Paper Award in the IEEE International Conference on Consumer Electronics-China, in 2014 and 2015.

ANDREW KWOK FAI LUI received the Ph.D. degree from The Australian National University, Canberra, ACT, Australia, in 1998. He is currently an Honorary Professor with the Department of Electronic Engineering and Computer Science, Hong Kong Metropolitan University. He recently joined the Queensland University of Technology as a Senior Research Engineer of the Research Engineering Facility. His current research interests include computational intelligence, autonomous systems, traffic modeling, and computer science education.