



OPEN

# A study on deep learning model based on global–local structure for crowd flow prediction

HeounMo Go<sup>1,2</sup> & SangHyun Park<sup>2</sup>✉

Crowd flow prediction has been studied for a variety of purposes, ranging from the private sector such as location selection of stores according to the characteristics of commercial districts and customer-tailored marketing to the public sector for social infrastructure design such as transportation networks. Its importance is even greater in light of the spread of contagious diseases such as COVID-19. In many cases, crowd flow can be divided into subgroups by common characteristics such as gender, age, location type, etc. If we use such hierarchical structure of the data effectively, we can improve prediction accuracy of crowd flow for subgroups. But the existing prediction models do not consider such hierarchical structure of the data. In this study, we propose a deep learning model based on global–local structure of the crowd flow data, which utilizes the overall(global) and subdivided by the types of sites(local) crowd flow data simultaneously to predict the crowd flow of each subgroup. The experiment result shows that the proposed model improves the prediction accuracy of each subdivided subgroup by 5.2% (Table 5 Cat #9)—45.95% (Table 11 Cat #5), depending on the data set. This result comes from the comparison with the related works under the same condition that use target category data to predict each subgroup. In addition, when we refine the global data composition by considering the correlation between subgroups and excluding low correlated subgroups, the prediction accuracy is further improved by 5.6–48.65%.

**Keywords** Crowd flow prediction, Deep learning, Spatio-temporal data mining

A significant amount of spatial-temporal data is being generated across many areas and the applications of such data are getting more important from target marketing to urban planning and public safety. In this paper, we focus on crowd flow density prediction, one of the spatial–temporal data prediction. Last year, in South Korea, there was a massive crowd crush in Seoul's Itaewon neighborhood and left many people dead and injured. If there was a public safety system based on crowd flow prediction, the tragedy could be avoided. As similar tragedies occur every year, the importance of crowd flow prediction is evident. As location-based services using smartphones become popular, research on crowd flow prediction using this data has also been actively conducted. In the private sector, it has been used for location selection of stores and customer-tailored marketing. In the public sector, it has been used for social infrastructure construction such as transportation networks. In public health area, with the recent spread of Covid-19, the crowd flow prediction is also gaining attention to prevent the spread of infectious diseases.

In many cases, the crowd flow data can be divided into subgroups. For example, the population can be divided by age and gender, and data from location based social network services can be also divided into subcategories by the type of visited spots such as food and shopping. In other words, The data exhibits hierarchical structure. To refine the effectiveness of crowd flow predictions, we need to account not only for the entire population but also for each sub-category. For example, for personalized marketing that is tailored to a specific demographic, it would be more effective to predict the movement of that demographic rather than the entire population. We propose a model structure that leverages the hierarchical structural characteristics of the data to maximize its utilization. To the best of our knowledge, our study is the first to concentrate on the hierarchical structure of crowd flow data.

<sup>1</sup>AI Service Division, SK Telecom, Eulji-ro 65, Jung-gu, Seoul 04539, Republic of Korea. <sup>2</sup>Department of Computer Science, Yonsei University, Yonsei-ro 50, Seodaemun-gu, Seoul 03722, Republic of Korea. ✉email: sanghyun@yonsei.ac.kr

## Related work

There are some previous works on predicting an individual's movement based on their location history<sup>1,2</sup>. These works focus on an individual, rather than for overall crowd-flow. Due to the growing popularity of deep learning, it has also been applied to crowd flow prediction. Deep learning-based models, such as Deep-ST<sup>3</sup>, which adapts a convolutional neural network<sup>4</sup> to model the spatial relations, has been proposed. In addition, ST-ResNet<sup>5</sup> has been suggested by using a residual framework<sup>6</sup> instead of the general convolution operation. DeepSTN+<sup>7</sup>, which considers the long-range spatial dependency more precisely, has been proposed.

On the other hand, RNN<sup>8</sup> based models have been suggested because crowd flow data have time-series nature. Models such as ConvLSTM<sup>9</sup>, hybrid approach<sup>10</sup>, STRCN<sup>11</sup> have been suggested with combining convolution and LSTM. In addition, Periodic-CRN<sup>12</sup>, which combines the pyramidal ConvGRU model with periodic representations to explicitly model the periodic nature of crowd flows, has been designed.

In addition, models specialized for traffic flow prediction in large cities, such as<sup>13,14</sup>, have been proposed. Particularly, novel methods of integrating inherent patterns in traffic flows, such as fixed spatial distributions and topological correlations<sup>15</sup>, and utilizing the GAN technique to predict traffic patterns with incomplete surveillance information<sup>16</sup>, have been proposed.

In the case of fine-grained data sets for metropolitan areas, a CNN-based model for enhancing computational efficiency, such as<sup>17</sup>, has been proposed. It converts the grid-based structure into a region-based one and models the flow between regions, and it shows promising prediction accuracy and computational performance. DeepCrowd<sup>18</sup> has been proposed, which adapts pyramid convLSTM and advanced attention mechanisms and is applied to fine-grained data sets. GridFormer<sup>19</sup>, which utilizes a periodically shifted sampling method and an attention mechanism to manage period disturbances, along with a pyramid 3D Swin transformers to capture long-range dependencies in a hierarchical fashion, has been proposed.

Additionally, research has been conducted on spatiotemporal multi-factor analysis, suggesting a new model architecture that learns features at all levels and captures spatiotemporal correlations at all-time stages using a modified densely connected convolutional network<sup>20</sup>

However, the existing models do not take into account the hierarchical structure of data described above. For example, when predicting one of the subcategories of data, existing models typically use only the target category data, without fully utilizing all the other categories of data. On the other hand, our novel model utilizes both the aggregated data (global) and the segmented data (local) according to the types of POI, consequently delivering improved performance compared to the existing models. Here, POI refers to Point of Interest, a location that a person visits. In addition, we propose a method for configuring global data that takes into account not only all categories but also the correlation between categories. This method of utilizing the hierarchical structure of global-local data has been previously used for high-dimensional multivariate time series data prediction<sup>21</sup>, but this is the first application of this approach to crowd flow prediction.

## Methods

### Problem formulation

To represent the regions of the city, we divide a city into an  $H \times W$  grids by the longitude and latitude, where all grids have the same size and each grid indicates a region. We define crowd flow density of a category for the region  $(h, w)$  at the  $i$ th time interval as follows:

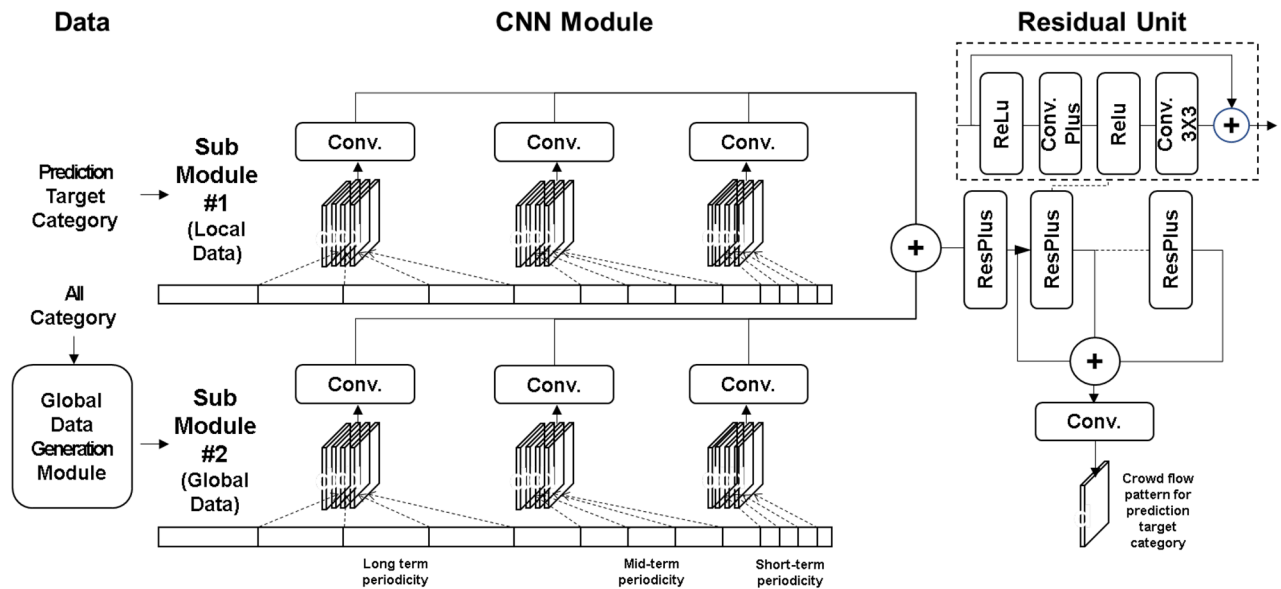
$$x_i^{h,w,c} = |\{j \geq 1 \mid g_j \in (h, w) \ \& \ c = \text{category}\}|$$

Here,  $g_j$  is the geospatial coordinate;  $g_j \in (h, w)$  means the point  $g_j$  lies within grid  $(h, w)$ , and vice versa;  $|\cdot|$  represents the cardinality of a set.  $c$  is the category of the point  $g_j$  belongs to. The target of crowd flow density prediction is to predict  $X_n$ , given the historical observations  $\{X_i \mid i = 1, 2, \dots, n-1\}$

### Model

In this study, we adopt DeepSTN+<sup>7</sup> as the base model. This is because it is a CNN-based model that can efficiently identify patterns on the grid structure and is among the most recent models for crowd flow prediction. However, the core idea of this study, which is the global-local structure approach, can be adapted to other models with some modifications.

Figure 1 shows the deep learning model proposed in this study based on global-local structure by leveraging the base model. First, input data is divided into local data, which is the target category data to be predicted, and global data, which includes all or some other categories. The CNN module is split into sub-module 1 for modeling local data and sub-module 2 for modeling global data. Local data is the input of CNN sub-module 1, and global data is the input of the global data generation module. The global data generation module is responsible for consisting global data by either including all categories as they are or removing some low correlated categories from the global data through calculating the correlation between the target category and each other category. Our assumption is that the lower a category is correlated with target category, the more excluding the category from global data improves the prediction performance. More specifically, when considering the correlation between categories, we select the top  $M$  categories with the highest correlation to the target category, and any categories with a correlation higher than  $N$  are also included. We call  $N$  correlation threshold. In this experiment,  $M$  and  $N$  are selected empirically based on the experimental results. In the case of Foursquare,  $M = 5$  and  $N = 0.31$  are selected because correlations vary significantly between categories, and in Gowalla,  $M = 8$  and  $N = 1$  are chosen as the correlations between categories are relatively high and the values are similar. We propose that, in instances where there are significant differences in the correlations between categories, it is advisable to opt for lower values of  $M$  and  $N$ . This is primarily because it is generally more beneficial to exclude categories



**Figure 1.** Model architecture.

with low correlations. Conversely, when the correlations between categories are high, it is preferable to choose higher values of  $M$  and  $N$ , as it is desirable to include as many highly correlated categories as possible. In this study, we use pearson correlation.

The subsequent CNN module is split into sub-module 1 for modeling local data and sub-module 2 for modeling global data. Each sub-module is composed of three CNN modules for modeling short-term periodicity, medium-term periodicity, and long-term periodicity. The residual unit module consists of repeated residual units. In each residual unit, the convolution process is repeated twice after the ReLU activation function, where the first convolution is to learn long-distance spatial patterns by non-adjacent filter coefficients. This is called as the Separated Channel structure, which is capable of modeling patterns between non-adjacent grids. The second convolution is a normal convolution for modeling patterns between adjacent grids. This results in the spatially predicted crowd flow pattern through the ReLU activation function, batch normalization<sup>22</sup>, Dropout<sup>23</sup>, CNN and tanh activation functions. We do not include the detailed POI distribution modeling component in our model. While this external factor can improve prediction accuracy for a specific area, it also makes it difficult to apply the model to areas without this information, thereby limiting its versatility. For instance, the Gowalla data used in this study does not contain detailed POI information. Moreover, in real-world scenarios, collecting detailed POI data can be time-consuming and expensive, and in some cases, impossible. In contrast, data from location-based services can generally be obtained more easily. For the aforementioned reasons, we propose that a model devoid of detailed POI information is more pragmatic from the perspective of a real-world business application.

The key contribution of this study lies in our recognition of the hierarchical structural characteristics of the data, leading us to propose a model structure that can maximize its utilization. This is achieved through the addition of 1) a global data generation module and 2) a sub-module 2 for integrating global data with local data. The internal structure of sub-module 2 mirrors that of sub-module 1, which is derived from the baseline model. However, we propose an optimal architecture for their combination. Experimental results indicate that the optimal point for merging the outputs of global and local data modeling is located between the output of the CNN module and the input of the Residual unit module. Consequently, the proposed model, as depicted in Fig. 1, demonstrates superior performance.

## Experiment

### Data

In this study, we use data from location-based social network services, Foursquare<sup>24</sup> and Gowalla<sup>25</sup>, which are suitable to test the proposed approach because they can be divided into subgroups by the types of visits. Furthermore, it is possible to conduct experiments by changing the scope of the target area because Gowalla data are collected all over the world.

As can be seen from the Table 1, Foursquare data are collected in Tokyo and New York for about 10 months from April 12, 2012 to February 16, 2013, and include check-in location and time, and ID and type of check-in location. The type of place is divided into a total of 10 sub-types: Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Service, Travel & Transport. 573,703 check-ins in Tokyo and 227,428 check-ins in New York are included. Gowalla data are collected for about 7 months from November 2010 to May 2011 and consists of 36,001,959 check-ins by 319,063 users in 2,844,076 locations around the world. They can be divided into 7 sub-types: Community, Entertainment, Food, Nightlife, Outdoors, Shopping, and Travel.

| Dataset    | Tokyo<br>(Foursquare) | New York<br>(Foursquare) | Global<br>(Gowalla) |
|------------|-----------------------|--------------------------|---------------------|
| Users      | 1939                  | 824                      | 319,063             |
| Venues     | 61,858                | 38,336                   | 2,844,076           |
| Check-ins  | 573,703               | 227,428                  | 36,001,959          |
| Categories | 10                    | 10                       | 7                   |

Table 1. Dataset statistics.

Preparation

To convert check-in data into grid structure-based crowd density data, the entire range of latitude and longitude of data set is divided into a number of grid, and the number of check-ins in each grid is calculated. In Foursquare Tokyo and New York, the entire coverage of the city is divided into 168 grids, resulting that each grid covers about 3 square kilometers which is approximately similar with the coverage of a government administrative region. The Gowalla data come from all around the world which is a wide range of area, and are further tested by dividing into two clusters, Gowalla America and Gowalla Asia/Europe/Africa, etc. K-means clustering is used for dividing entire data set into two clusters. Two types of global data are consisted, one includes all categories, while the other excludes low correlated categories. The prediction performance when using those two global data is compared.

Training

The ratio of the training data to the test data is 3:1. Mean Square Error is used as the loss function, and Root Mean Square Error is adopted as the accuracy evaluation metric. In this experiment, the most commonly used RMSE index is used because it focuses on comparing the prediction performance between models. We update the weights through the Adam optimization algorithm. The dropout rate is 0.1, and the number of residual units is set to 10. The learning rate is  $10^{-4}$ , the number of epochs is 800, and the batch size is 32.

Inference

The final performance is compared by using the test data which is not used for training. The final RMSE is calculated by averaging the RMSEs except for the maximum/minimum RMSE and by repeating the experiment 10 times using the model trained with the training data. Root mean square error is used as performance metric, and RMSE difference of less than 5% is judged to be equivalent. This study conducted performance comparison research with a baseline model that does not include the proposed techniques, in order to validate the effects of proposed model. Additionally, as part of the ablation experiment, comparative experiments were conducted considering and not considering correlation when generating global data. These results are described in the following section. Also, the qualitative analysis of the experimental results is described in the following section.

Result

In the RMSE comparison table, 'Model (A)' uses global data composed of all other categories, while 'Model (B)' uses global data consisting only of highly correlated categories with the target category, which is marked in bold in the correlation table.

Foursquare Tokyo

Table 2 presents the correlation of data across categories, which significantly impacts the performance of the proposed model. Table 3 illustrates the prediction accuracy of both the baseline and proposed models. The

| Cat. | #1           | #2           | #3     | #4           | #5    | #6           | #7           | #8     | #9           | #10          |
|------|--------------|--------------|--------|--------------|-------|--------------|--------------|--------|--------------|--------------|
| #1   | 1.000        | <b>0.335</b> | 0.165  | <b>0.476</b> | 0.150 | <b>0.311</b> | <b>0.330</b> | 0.131  | <b>0.462</b> | <b>0.479</b> |
| #2   | 0.335        | 1.000        | 0.134  | 0.272        | 0.080 | 0.168        | 0.208        | 0.073  | 0.285        | 0.303        |
| #3   | 0.165        | 0.134        | 1.000  | 0.241        | 0.093 | 0.124        | 0.112        | -0.008 | 0.121        | 0.242        |
| #4   | <b>0.476</b> | 0.272        | 0.241  | 1.000        | 0.197 | <b>0.428</b> | <b>0.445</b> | 0.153  | <b>0.596</b> | <b>0.651</b> |
| #5   | 0.150        | 0.080        | 0.093  | 0.197        | 1.000 | 0.128        | 0.121        | 0.064  | 0.171        | 0.201        |
| #6   | 0.311        | 0.168        | 0.124  | 0.428        | 0.128 | 1.000        | 0.381        | 0.113  | 0.415        | 0.413        |
| #7   | <b>0.330</b> | 0.208        | 0.112  | <b>0.445</b> | 0.121 | <b>0.381</b> | 1.000        | 0.099  | <b>0.430</b> | <b>0.442</b> |
| #8   | 0.131        | 0.073        | -0.008 | 0.153        | 0.064 | 0.113        | 0.099        | 1.000  | 0.153        | 0.161        |
| #9   | <b>0.462</b> | 0.285        | 0.121  | <b>0.596</b> | 0.171 | <b>0.415</b> | <b>0.430</b> | 0.153  | 1.000        | <b>0.605</b> |
| #10  | <b>0.479</b> | 0.303        | 0.242  | <b>0.651</b> | 0.201 | <b>0.413</b> | <b>0.442</b> | 0.161  | <b>0.605</b> | 1.000        |

Table 2. Foursquare Tokyo—Correlation between categories.

| Cat. | STResNet | DeepSTN+ | Model(A)     | RMSE ↓      | Model(B)      | RMSE ↓       |
|------|----------|----------|--------------|-------------|---------------|--------------|
| #1   | 0.0479   | 0.0406   | 0.0384       | 5.4%        | <b>0.0379</b> | <b>6.7%</b>  |
| #2   | 0.0289   | 0.0274   | 0.0269       | –           | 0.0269        | –            |
| #3   | 0.0146   | 0.0146   | 0.0146       | –           | 0.0146        | –            |
| #4   | 0.0581   | 0.0409   | 0.0384       | 6.1%        | <b>0.038</b>  | <b>7.1%</b>  |
| #5   | 0.0409   | 0.0408   | 0.0409       | –           | 0.0409        | –            |
| #6   | 0.0107   | 0.01     | 0.01         | –           | 0.01          | –            |
| #7   | 0.0412   | 0.036    | <b>0.034</b> | <b>5.6%</b> | <b>0.034</b>  | <b>5.6%</b>  |
| #8   | 0.042    | 0.042    | 0.042        | –           | 0.042         | –            |
| #9   | 0.0357   | 0.0285   | 0.0255       | 10.5%       | <b>0.0254</b> | <b>10.9%</b> |
| #10  | 0.0667   | 0.0449   | <b>0.043</b> | <b>6.9%</b> | <b>0.043</b>  | <b>6.9%</b>  |

**Table 3.** Foursquare Tokyo—RMSE comparison.

performance improvement of the proposed method over other methods can be attributed to its unique approach. To predict one of the data subcategories, unlike existing models that typically rely solely on the target category data, our method employs a combination of both aggregated (global) and segmented (local) data.

From Table 2, it can be observed that the overall correlation ranges from  $-0.008$  to  $0.651$ . The correlation between categories, on the whole, is not particularly strong, with the highest correlation being  $0.651$ . This is also reflected in the RMSE comparison results in Table 3, which demonstrate that the performance improvement is not substantial when using global data. Specifically, when the correlation between categories is higher, the prediction performance improves more significantly by incorporating global data.

Category 1 (Arts & Entertainment) exhibits a correlation of  $0.131 - 0.479$  with other categories, and it is observed that 6 categories have correlations higher than the correlation threshold of  $N(0.31)$ , which is relatively high in this case. This correlation pattern results in a  $5.4\%$  reduction in error when all categories are included as global data. Furthermore, when experiments are conducted using global data but categories 3, 5, and 8 are excluded, which have relatively low correlation with the target category, an additional  $1.3\%$  reduction in error is observed in Table 3. It can be seen that category 4 (Food) and category 9 (Shop & Service) also exhibit a similar trend.

Regarding Category 7 (Professional & Other Places) and Category 10 (Travel & Transport), it is observed that the prediction error is reduced when configuring global data with all categories, but no additional improvement is observed when global data is configured based on correlation. This requires further analysis.

In the case of categories 2 (College & University), 3 (Event), 5 (Nightlife Spot), 6 (Outdoors & Recreation), and 8 (Outdoors & Recreation), it appears that there is no performance improvement when using global data because the correlation between these categories and other categories is lower than the threshold value of  $N$ .

Figure 2 shows that the proposed model provides better performance than base models in Foursquare Tokyo dataset.

### Foursquare New York

Table 4 shows that the correlation between categories ranges from  $-0.008$  to  $0.655$ . Overall, the correlation between categories is not particularly strong, with the maximum correlation being  $0.655$ , which is also reflected in the RMSE performance results in Table 5. The performance improvement ranges from  $5.2$  to  $6.96\%$ , as shown in Table 5, which is less significant compared to the performance gain observed in the Gowalla data, as discussed later. In addition, the performance improvement tends to be proportional to the correlation between categories. Category 1 (Arts & Entertainment) shows a correlation of  $0.086$  to  $0.589$ , and 5 categories have a significant correlation with Category 1 higher than the threshold of  $N(0.31)$ .

This results in a  $6.1\%$  reduction in error when all categories are included as global data. Furthermore, when configuring the global data by excluding categories 2, 3, 5, and 8, which have relatively low correlation, an additional  $0.7\%$  reduction in error is achieved.

Categories 6 (Outdoors & Recreation), 7 (Professional), and 9 (Shop & Service) also exhibit similar trends.

Regarding Category 4 (Food), while including all categories in global data results in a  $6.4\%$  reduction in errors, no further improvement is observed when global data is configured with consideration of correlation. Similarly, for Category 10 (Travel & Transport), no performance gain is observed when using global data, and this requires further investigation.

### Gowalla global

Looking at the correlation analysis results between categories in Table 6, it can be seen that the overall correlations range from a minimum of  $0.775$  to a maximum of  $0.906$ . Considering that the correlation between categories in Foursquare data is up to  $0.65$ , we can conclude that the correlations between categories in Gowalla are generally stronger than those in Foursquare. Looking at the experimental results in Table 7, the error reduction for Gowalla ranges from  $6.3$  to  $36.1\%$ , which is significantly higher than that of Foursquare, which ranges from  $\sim 10.9\%$ . Once again, we see that the performance improvement when using global data tends to be proportional to the correlation between categories.

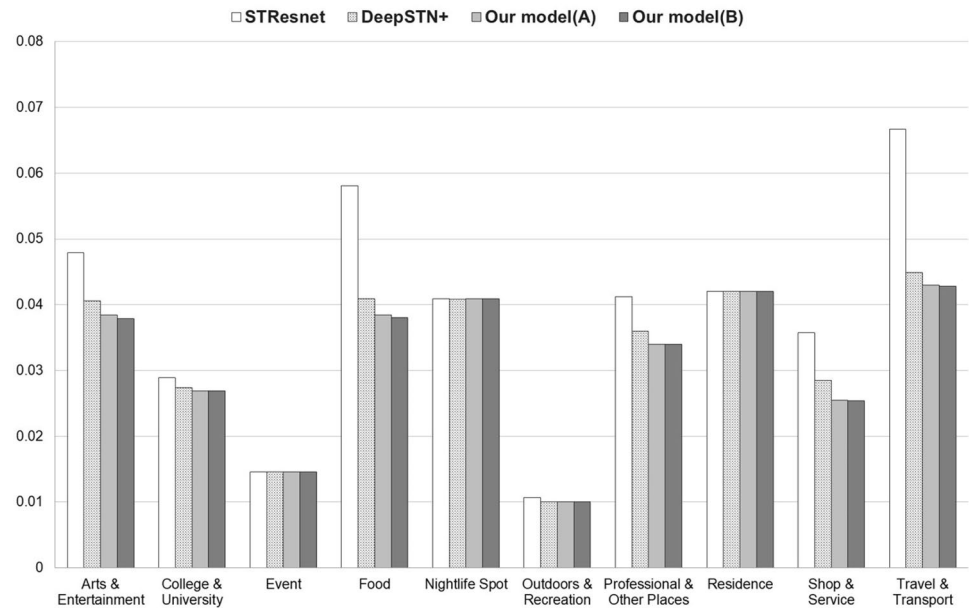


Figure 2. Foursquare Tokyo RMSE comparison.

| Cat. | #1           | #2    | #3     | #4           | #5    | #6           | #7           | #8      | #9           | #10          |
|------|--------------|-------|--------|--------------|-------|--------------|--------------|---------|--------------|--------------|
| #1   | 1.000        | 0.282 | 0.202  | <b>0.589</b> | 0.195 | <b>0.444</b> | <b>0.400</b> | 0.086   | <b>0.474</b> | <b>0.427</b> |
| #2   | 0.282        | 1.000 | 0.105  | 0.283        | 0.078 | 0.236        | 0.242        | 0.052   | 0.254        | 0.212        |
| #3   | 0.202        | 0.105 | 1.000  | 0.245        | 0.068 | 0.174        | 0.180        | − 0.008 | 0.199        | 0.159        |
| #4   | <b>0.589</b> | 0.283 | 0.245  | 1.000        | 0.219 | <b>0.601</b> | <b>0.577</b> | 0.124   | <b>0.655</b> | <b>0.567</b> |
| #5   | 0.195        | 0.078 | 0.068  | 0.219        | 1.000 | 0.188        | 0.151        | 0.016   | 0.191        | 0.201        |
| #6   | <b>0.444</b> | 0.236 | 0.174  | <b>0.601</b> | 0.188 | 1.000        | <b>0.494</b> | 0.102   | <b>0.524</b> | <b>0.495</b> |
| #7   | <b>0.400</b> | 0.242 | 0.180  | <b>0.577</b> | 0.151 | <b>0.494</b> | 1.000        | 0.105   | <b>0.514</b> | <b>0.454</b> |
| #8   | 0.086        | 0.052 | −0.008 | 0.124        | 0.016 | 0.102        | 0.105        | 1.000   | 0.118        | 0.087        |
| #9   | <b>0.474</b> | 0.254 | 0.199  | <b>0.655</b> | 0.191 | <b>0.524</b> | <b>0.514</b> | 0.118   | 1.000        | <b>0.507</b> |
| #10  | <b>0.427</b> | 0.212 | 0.159  | <b>0.567</b> | 0.201 | <b>0.495</b> | <b>0.454</b> | 0.087   | <b>0.507</b> | 1.000        |

Table 4. Foursquare New York—Correlation between categories.

| Cat. | STResNet | DeepSTN+ | Model(A)      | RMSE ↓      | Model(B)      | RMSE ↓       |
|------|----------|----------|---------------|-------------|---------------|--------------|
| #1   | 0.0163   | 0.0147   | 0.0138        | 6.1%        | <b>0.0136</b> | <b>6.8%</b>  |
| #2   | 0.029    | 0.0283   | 0.0283        | –           | 0.0283        | –            |
| #3   | 0.0152   | 0.0152   | 0.0152        | –           | 0.0152        | –            |
| #4   | 0.0301   | 0.0205   | <b>0.0192</b> | <b>6.4%</b> | <b>0.0192</b> | <b>6.4%</b>  |
| #5   | 0.0063   | 0.0063   | 0.0063        | –           | 0.0063        | –            |
| #6   | 0.0178   | 0.0158   | 0.0149        | 5.7%        | <b>0.0147</b> | <b>6.96%</b> |
| #7   | 0.0229   | 0.0192   | <b>0.0182</b> | <b>5.3%</b> | <b>0.018</b>  | <b>6.25%</b> |
| #8   | 0.0253   | 0.0253   | 0.0252        | –           | 0.0252        | –            |
| #9   | 0.0267   | 0.0191   | 0.0181        | 5.2%        | <b>0.0179</b> | <b>6.3%</b>  |
| #10  | 0.0209   | 0.0166   | <b>0.0164</b> | –           | <b>0.0165</b> | –            |

Table 5. Foursquare New York—RMSE comparison.



| Category         | #1           | #2           | #3           | #4           | #5           | #6           | #7           |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| #1_Community     | 1.000        | 0.838        | <b>0.844</b> | 0.818        | <b>0.856</b> | <b>0.871</b> | <b>0.852</b> |
| #2_Entertainment | <b>0.838</b> | 1.000        | 0.816        | <b>0.818</b> | <b>0.854</b> | <b>0.865</b> | 0.815        |
| #3_Food          | <b>0.844</b> | <b>0.816</b> | 1.000        | 0.802        | <b>0.809</b> | <b>0.906</b> | 0.775        |
| #4_Nightlife     | <b>0.818</b> | <b>0.818</b> | 0.802        | 1.000        | <b>0.806</b> | <b>0.840</b> | 0.778        |
| #5_Outdoor       | 0.856        | 0.854        | 0.809        | 0.806        | 1.000        | 0.845        | 0.815        |
| #6_Shopping      | 0.871        | 0.865        | 0.906        | 0.840        | 0.845        | 1.000        | 0.824        |
| #7_Travel        | <b>0.852</b> | <b>0.815</b> | 0.775        | 0.778        | <b>0.815</b> | <b>0.824</b> | 1.000        |

**Table 6.** Gowalla Global—Correlation between categories.

| Cat. | STResNet | DeepSTN+ | Model(A)      | RMSE ↓        | Model(B)      | RMSE ↓        |
|------|----------|----------|---------------|---------------|---------------|---------------|
| #1   | 0.0293   | 0.0135   | <b>0.012</b>  | <b>11.11%</b> | <b>0.012</b>  | <b>11.11%</b> |
| #2   | 0.0091   | 0.008    | <b>0.0075</b> | <b>6.3%</b>   | <b>0.0075</b> | <b>6.3%</b>   |
| #3   | 0.0173   | 0.0119   | 0.0107        | 10.1%         | <b>0.0098</b> | <b>17.65%</b> |
| #4   | 0.0065   | 0.0054   | <b>0.005</b>  | <b>7.4%</b>   | <b>0.005</b>  | <b>7.4%</b>   |
| #5   | 0.0133   | 0.0097   | 0.0095        | –             | 0.0095        | –             |
| #6   | 0.0148   | 0.0077   | 0.0074        | –             | 0.0074        | –             |
| #7   | 0.0101   | 0.0097   | 0.0066        | 32%           | <b>0.0062</b> | <b>36.1%</b>  |

**Table 7.** Gowalla Global—RMSE comparison.

Additionally, the differences in correlation between categories in Gowalla are smaller than those in Foursquare. Thus, we can expect that the performance improvement when global data is configured considering correlation would be smaller in Gowalla than in Foursquare.

When composing global data by considering correlation in Foursquare, there are 5 categories that show performance improvement. However, in Gowalla, only 2 categories show additional performance improvements. We can infer that as the difference in correlation between categories increases, the performance tends to be better when composing global data considering correlation.

In the case of category 3 (Food), the correlation with other categories ranges from 0.775 to 0.906, and the difference between the maximum and minimum values of correlation is 0.131, which is relatively high compared to other categories in Gowalla global. It can be expected that there will be additional performance improvement when global data is configured considering correlation. As expected, the experiment results show that the prediction error is 7.55% lower when global data is configured considering correlation. Category 7 (Travel) also shows a similar trend.

In categories 1 (Community), 2 (Entertainment), and 4 (Nightlife), the correlation between categories is high, and the difference in correlation between categories is also relatively small. This may explain why there is no additional performance improvement observed when configuring global data with consideration of correlation in these categories.

Figure 3 shows that the proposed model provides better performance than base models in Gowalla Global dataset.

### Gowalla America

Looking at the correlation analysis results between categories in Table 8, it can be seen that the overall correlations range from 0.797 to 0.9. From the experimental results in Table 9, the performance improvement is from 12.5 to 43.5%, which is higher than the performance improvement of Gowalla global. The reason for this, regardless of no significant difference in the correlations, would be inferred from the amount of data. In the case of Gowalla America, the number of samples is smaller because only samples from the Americas region are selected from Gowalla global. It can be seen that the smaller the number of samples, the better the performance improvement of using global data including all the other categories.

In Gowalla America data, the correlations between the target measurement category and other categories are high overall, and the differences in the correlation between categories are small. Therefore, when global data is composed of all categories, the performance improves significantly, whereas when composing global data considering correlation, additional performance improvement is not significant.

This can be confirmed from the experimental results in Table 9. When global data is composed of all categories, the average error reduction is 22%, which is 8.6% higher than that of Gowalla Global (13.4%). When the global data is configured in consideration of correlation, there is a slight improvement, but the RMSEs themselves are not significantly different.

Figure 4 shows that the proposed model provides better performance than base models in Gowalla America dataset.

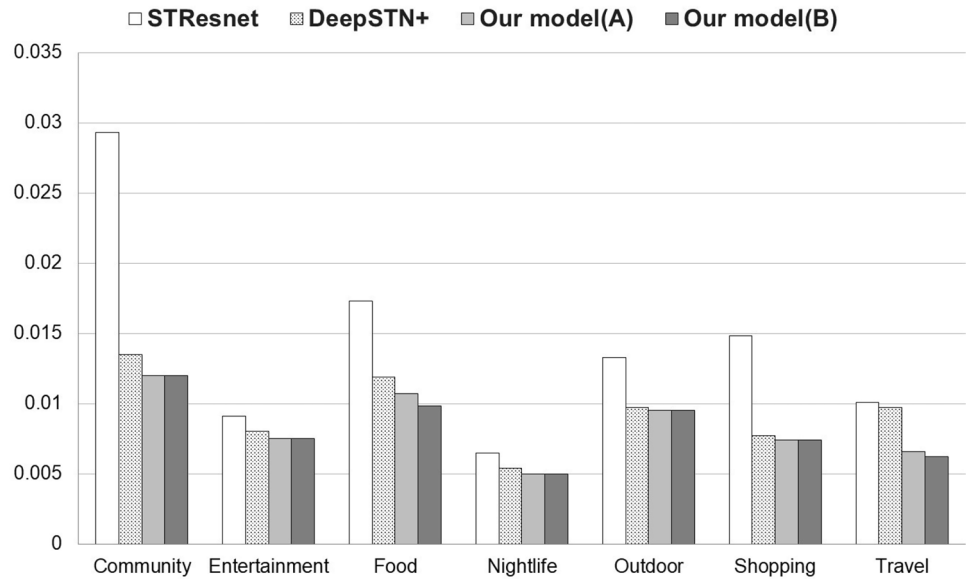


Figure 3. Gowalla global RMSE comparison.

| Category         | #1           | #2           | #3           | #4           | #5           | #6           | #7           |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| #1_Community     | 1.000        | <b>0.900</b> | 0.838        | 0.833        | <b>0.875</b> | <b>0.881</b> | <b>0.856</b> |
| #2_Entertainment | <b>0.900</b> | 1.000        | 0.803        | <b>0.846</b> | <b>0.885</b> | <b>0.856</b> | 0.844        |
| #3_Food          | <b>0.838</b> | 0.803        | 1.000        | <b>0.810</b> | 0.797        | <b>0.869</b> | <b>0.829</b> |
| #4_Nightlife     | <b>0.833</b> | <b>0.846</b> | 0.810        | 1.000        | <b>0.811</b> | <b>0.839</b> | 0.802        |
| #5_Outdoor       | <b>0.875</b> | <b>0.885</b> | 0.797        | 0.811        | 1.000        | <b>0.844</b> | <b>0.837</b> |
| #6_Shopping      | <b>0.881</b> | <b>0.856</b> | <b>0.869</b> | <b>0.839</b> | <b>0.844</b> | 1.000        | <b>0.853</b> |
| #7_Travel        | <b>0.856</b> | <b>0.844</b> | 0.829        | 0.802        | <b>0.837</b> | <b>0.853</b> | 1.000        |

Table 8. Gowalla America—Correlation between categories.

| Cat. | STResNet | DeepSTN+ | Model(A)      | RMSE ↓       | Model(B)      | RMSE ↓        |
|------|----------|----------|---------------|--------------|---------------|---------------|
| #1   | 0.0245   | 0.0177   | 0.011         | 37.85%       | <b>0.01</b>   | <b>43.50%</b> |
| #2   | 0.0093   | 0.0087   | 0.0071        | 18.4%        | <b>0.0064</b> | <b>26.4%</b>  |
| #3   | 0.0134   | 0.0123   | 0.01          | 18.70%       | <b>0.009</b>  | <b>26.83%</b> |
| #4   | 0.0091   | 0.0088   | 0.0062        | 29.5%        | <b>0.006</b>  | <b>31.8%</b>  |
| #5   | 0.0235   | 0.0159   | 0.013         | 18.24%       | <b>0.0129</b> | <b>18.24%</b> |
| #6   | 0.0132   | 0.0128   | 0.0104        | 18.75%       | <b>0.010</b>  | <b>18.75%</b> |
| #7   | 0.0106   | 0.0096   | <b>0.0084</b> | <b>12.5%</b> | <b>0.0084</b> | <b>12.5%</b>  |

Table 9. Gowalla America—RMSE comparison.

Gowalla Asia/Europe/Africa, etc

Looking at the correlation analysis results between categories in Table 10, it can be seen that the overall correlations range from 0.734–0.921. According to the experimental results in Table 11, the RMSE reduction of Gowalla Asia/Europe/Africa is from 9.71 to 48.65%, which is similar to the performance improvement level of Gowalla America.

Figure 5 shows that the proposed model provides better performance than base models in Asia/Europe/Africa dataset.

Conclusion

In this study, we propose a deep learning model that takes into account the global-local structure of crowd flow data. The model utilizes both overall (global) and subdivided (local) crowd flow data simultaneously, with the goal of increasing the prediction accuracy for each sub-divided subgroup. Furthermore, with respect to global



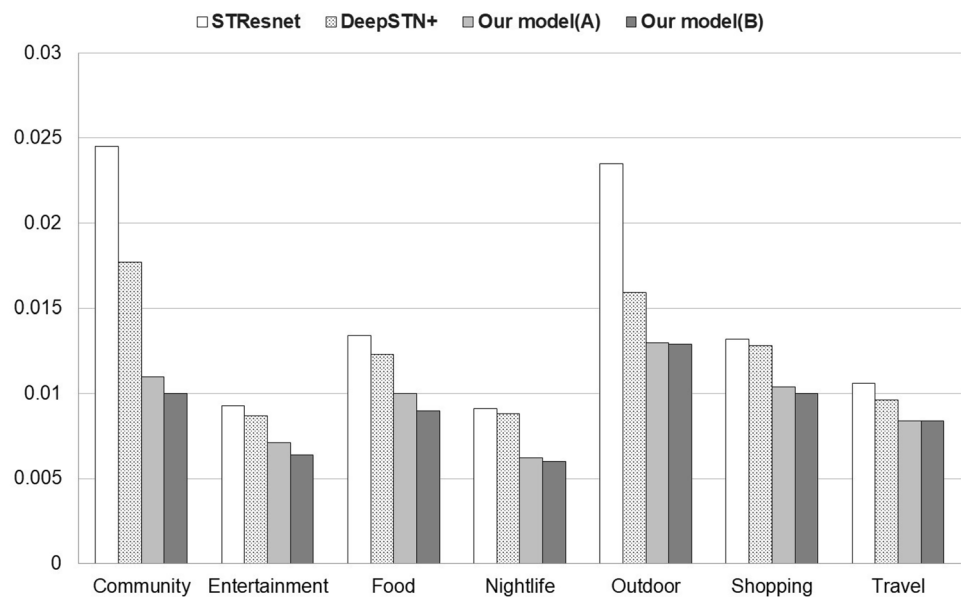


Figure 4. Gowalla America RMSE comparison.

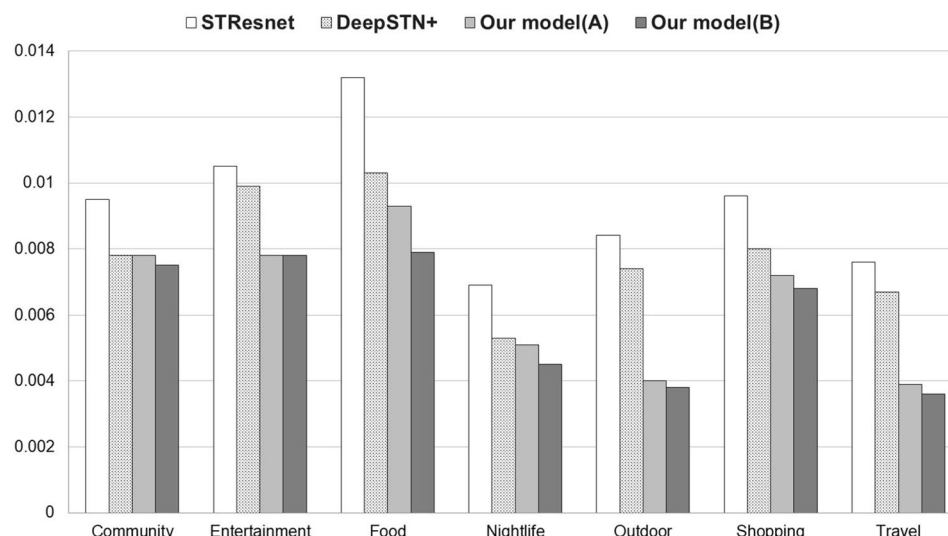
| Category         | #1           | #2           | #3           | #4           | #5           | #6           | #7    |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|-------|
| #1_Community     | 1.000        | 0.791        | 0.857        | 0.782        | <b>0.801</b> | <b>0.866</b> | 0.756 |
| #2_Entertainment | <b>0.791</b> | 1.000        | <b>0.807</b> | 0.740        | <b>0.795</b> | <b>0.799</b> | 0.745 |
| #3_Food          | <b>0.857</b> | 0.807        | 1.000        | <b>0.822</b> | <b>0.822</b> | <b>0.921</b> | 0.778 |
| #4_Nightlife     | <b>0.782</b> | <b>0.740</b> | <b>0.822</b> | 1.000        | 0.739        | <b>0.817</b> | 0.734 |
| #5_Outdoor       | <b>0.801</b> | <b>0.795</b> | <b>0.822</b> | 0.739        | 1.000        | <b>0.802</b> | 0.755 |
| #6_Shopping      | <b>0.866</b> | 0.799        | <b>0.921</b> | <b>0.921</b> | <b>0.802</b> | 1.000        | 0.770 |
| #7_Travel        | <b>0.756</b> | 0.745        | <b>0.778</b> | 0.734        | <b>0.755</b> | <b>0.770</b> | 1.000 |

Table 10. Gowalla Asia/Europe/Africa, etc.—Correlation between categories.

| Cat. | STResNet | DeepSTN+ | Model(A)      | RMSE ↓        | Model(B)      | RMSE ↓        |
|------|----------|----------|---------------|---------------|---------------|---------------|
| #1   | 0.0095   | 0.0078   | 0.0078        | —             | 0.0075        | —             |
| #2   | 0.0105   | 0.0099   | <b>0.0078</b> | <b>21.21%</b> | <b>0.0078</b> | <b>21.21%</b> |
| #3   | 0.0132   | 0.0103   | 0.0093        | 9.71%         | <b>0.0079</b> | <b>23.30%</b> |
| #4   | 0.0069   | 0.0053   | 0.0051        | —             | <b>0.0045</b> | <b>15.09%</b> |
| #5   | 0.0084   | 0.0074   | 0.004         | 45.95%        | <b>0.0038</b> | <b>48.65%</b> |
| #6   | 0.0096   | 0.008    | 0.0072        | 10.00%        | <b>0.0068</b> | <b>15.00%</b> |
| #7   | 0.0076   | 0.0067   | 0.0039        | 41.79%        | <b>0.0036</b> | <b>46.27%</b> |

Table 11. Gowalla Asia/Europe/Africa, etc—RMSE comparison.

data, we propose a method that takes into account not only all categories but also the correlation between the prediction target category and each other category. Specifically, we exclude categories that have a low correlation with the target category from global data. The experiment result shows that the proposed model improves the prediction accuracy of each sub-divided subgroup by 5.2% (Table 5. Cat #9)—45.95% (Table 11. Cat #5), depending on the data set. This result comes from the comparison with the related works under the same condition that use target category data to predict each subgroup. In addition, when we advance the global data composition by considering the correlation between subgroups, which excludes low correlated subgroups from global data, the prediction accuracy is even more improved by 5.6% – 48.65%. Overall, if the correlation between the prediction category and other categories is high, the approach of combining local and global data shows high performance improvement. In principle, the entire crowd flow(global) data should be correlated with each category to some extent, which means that the proposed approach can be generalized to all kinds of crowd flow data. Furthermore,



**Figure 5.** Gowalla Asia/Europe/Africa, etc RMSE comparison.

the overall performance is enhanced when the global data is configured by considering correlation, which is excluding categories with relatively low correlation. The approach of excluding categories with low correlation to the target category from the global data tends to improve prediction performance more as the correlation between the target category and other categories decreases. This study can be utilized in various purpose such as tailored marketing, new drug development and so on.

## Data availability

Our code for the experiments is available at [https://github.com/ilyhs79/Glocal\\_STN](https://github.com/ilyhs79/Glocal_STN).

Received: 27 January 2024; Accepted: 27 May 2024

Published online: 01 June 2024

## References

1. Fan, Z., Song, X., Shibasaki, R. & Adachi, R. *Citymomentum: an online approach for crowd behavior prediction at a citywide level*. In ACM UbiComp, 559–569. ACM. (2015).
2. Song, X., Zhang, Q., Sekimoto, Y., Shibasaki, R. *Prediction of human emergency behavior and their mobility following large-scale disaster*. In ACM SIGKDD, 5–14. ACM. (2014).
3. Zhang, J., Zheng, Y., Qi, D., Li, R., & Yi, X. *Dnn based prediction model for spatio-temporal data*. Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 92. ACM. (2016).
4. Krizhevsky, A., Sutskever, I. & Hinton, G. E. *ImageNet classification with deep convolutional neural networks*. In NIPS (2012).
5. Zhang, et al. *Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction* AAAI 2017 (2017).
6. He, K., Zhang, X., Ren, S. & Sun, J. *Deep residual learning for image recognition*. In IEEE CVPR (2015).
7. Lin. *DeepSTN+ : Context-aware spatial-temporal neural network for crowd flow prediction* AAAI 2019 (2019).
8. Sutskever, I., Vinyals, O. & Le, Q. V. *Sequence to sequence learning with neural networks*. NIPS 27, 3104–3112 (2014).
9. Shi, X. et al. *Convolutional LSTM network: A machine learning approach for precipitation nowcasting*. arXiv:1506.04214 (2015).
10. Du, S., Li, T., Gong, X., Yang, Y. & Horng, S. J. *Traffic flow forecasting based on hybrid deep learning framework*. In Intelligent Systems and Knowledge Engineering (ISKE) 2017 12th International Conference on, 1–6. IEEE (2017).
11. Jin, W., Lin, Y., Wu, Z. & Wan, H. *Spatio-temporal recurrent convolutional networks for citywide short-term crowd flows prediction*. In Proceedings of the 2nd International Conference on Compute and Data Analysis, 28–35. ACM. (2018).
12. Zonoozi, A. et al. *A convolutional recurrent model for crowd density prediction with recurring periodic patterns* IJCAI. Periodic-CRN 2018, 3732–3738 (2018).
13. Zheng, C., et al. *DeepSTD: mining spatio-temporal disturbances of multiple context factors for citywide traffic flow prediction* TITS 2019. Paper (2019).
14. Zhang, J., et al. *Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit*. IET Intell. Transp. Syst. (2020).
15. Liu, M. et al. *Spatio-temporal autoencoder for traffic flow prediction*. IEEE TITS (2023).
16. Wang, Pengkun et al. *An ST-GAN method IEEE TVT, inferring intersection traffic patterns with sparse video surveillance information*, (2022).
17. Liang, Y. et al. *Fine-grained urban flow prediction*. WWW (2021).
18. Jiang, R., et al. *DeepCrowd: A deep model for large-scale citywide crowd density and flow prediction*. TKDE (2021).
19. Chaoqun, S. et al. *Spatial-temporal transformer network for citywide crowd flow prediction* ECAI, GridFormer, (2023).
20. Wang, P. et al. *Joint gated co-attention based multi-modal networks for subregion house price prediction*. IEEE TKDE 35, 1667 (2023).
21. Sen, R. *Think globally, act locally: a deep neural network approach to high-dimensional time series forecasting* arXiv 2019, Paper (2019).
22. Ioffe, S. & Szegedy, C. *Batch normalization: accelerating deep network training by reducing internal covariate shift*. arXiv preprint arXiv:1502.03167 (2015).
23. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. *Dropout: A simple way to prevent neural networks from overfitting*. J. Mach. Learn. Res. 15(1), 1929–1958 (2014).

24. Yang, D. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE TSMC* **45**, 129–142 (2015).
25. Liu, Y. Exploiting geographical neighborhood characteristics for location recommendation. *CIKM*, p 739–748 (2014).

## Acknowledgements

This research was supported by the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2023-00229822).

## Author contributions

HM.GO wrote the main manuscript text. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to S.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024