Optimizing systemic redundancy of traffic sensor networks while maintaining resilience: New evidence from using graph learning

Junqing Tang[1,2], Shufen Wei[1], , Xiaowei Li[3], Duo Li[4], *Senior Member IEEE*

***Abstract* – The optimization of systemic redundancy by minimizing the sensor quantity can improve the efficiency of sensor networks and save costs. However, from the perspective of risk management, this redundancy reduction can also bring a significant loss in the overall network resilience because the less the systemic redundancy is, the fewer backup components in the network when shocks hit and, therefore, the less overall resilience. In this paper, we investigate this intractable dilemma and attempt to pinpoint the tradeoff point for a city-scale automatic number plate recognition (ANPR) system in Cambridge, UK. By developing a two-stage graph deep learning (GDL) model, we first optimize the layout of the ANPR system to reduce redundancy and find its efficiency profile. Next, we study what effects this redundancy reduction can bring to the overall resilience, as the overall observability drops with the reduction in the number of sensors and find an optimal balance. The results show that our approach can effectively optimize the system's redundancy by using only 47% of the original sensors to reconstruct the full picture with a mean absolute error (MAE) of only 11.18 and a root mean square error (RMSE) of 19.49; most importantly, the overall system resilience is maintained at 70% in the meantime. This paper provides an alternative perspective for dealing with the well-known 'efficiency-resilience' dilemma and offers new evidence to enable better decision and policy making for city managers and planners in local authorities.**

***Index Terms*—Resilience; Sensor network; Optimization; Graph learning; ANPR**

I. Introduction

DIGITAL tools have been widely deployed in various smart city initiatives, such as smart grids and Internet of things (IoT) devices, to improve their operational efficiency, resource allocation, infrastructure management, and more [1]. In the transportation sector, digital technologies and tools have brought novel opportunities, offering new solutions to conventional traffic-related challenges [2]. Among them, one traffic sensing technology, namely the camera-based traffic monitoring technology, has been particularly popular in many countries and has been widely utilized to manage urban mobility and city planning.

For city-scale traffic monitoring, traffic sensors are normally implemented as fixed sensor arrays, forming a networked large-scale traffic monitoring system [3]. Some representative examples of camera-based traffic sensors include automatic number plate recognition (ANPR) sensors for determining vehicle identities, vehicle presence sensors (FLIR TrafiCam) for managing traffic flows at signalized junctions, and dual-vision thermal traffic cameras (ITS-Series Dual AID) for automatic incident detection. As a sensing technology associated with considerable cost, geographical layout optmization for sensor networks has long been an eye-catching research topic. Scholars have attempted to develop multiple approaches for tackling redundancy optimization in sensing networks to minimize the required sensor quantity and improve their overall efficiency. Network-wide observability plays an indispensable role in evaluating the effectiveness of a deployed sensor set, which depicts its power to capture vehicle fleets and the vehicle movements, i.e., how well the sensor network can perform in terms of traffic monitoring [4], [5]. With the rapid development of machine learning methods, this growing trend in sensor layout optimization become even more popular in recent years.

Parallel to layout optimization and efficiency improvement, one particular issue is also attracting scholars' attention: that is, the network resilience against component failures [6]. In resilience engineering theory, a system's resilience against disruptions can be enhanced by improving its capability to cope with component failures, where a certain level of systemic redundancy is a must [7], [8]. In other words, systemic redundancy is a prerequisite and plays a determinant role in maintaining system resilience. This leads to the well-known dilemma of balancing system efficiency (reducing redundancy) and system resilience (maintaining redundancy). For each side of this 'leverage', a considerable number of previous studies have been dedicated to investigating the issue in a unilateral fashion. However, little exploration has been performed regarding finding the aforementioned tradeoffs between optimizing the systemic redundancy of sensor networks while simultaneously maintaining their resilience.

To fill this gap, we devise a two-stage graph deep learning (GDL) model consisting of self-attention graph pooling (SAGPool) and a graph convolutional neural network (GCNN) to perform node selection and estimation. Taking a city-scale ANPR sensing network from Cambridge, UK, as an illustrative example, we first quantitatively demonstrate the capability of this developed model in terms of optimizing redundant sensors in an array. The model is then compared with other approaches, and its sensitivity and comparative merits are also verified. Two error measures are adopted, namely, the mean absolute error (MAE) and the root mean

[1]School of Urban Planning and Design, Shenzhen Graduate Schoole, Peking University, Shenzhen, 518055, China
[2]Key Laboratory of Earth Surface System and Human-earth Relations of Ministry of Natural Resources of China, Shenzhen Graduate School, Peking University, Shenzhen, 518055, China
[3]School of Civil Engineering, Xi'an University of Architecture & Technology, Xi'an 710055, China
[4]Department of Engineering, Nottingham Trent University, Nottingham NG1 4FQ, United Kingdom
Corresponding author: Duo Li (email: duo.li@ntu.ac.uk)

square error (RMSE). Next, after optimizing the redundancy, we assess the system's resilience in each optimization scenario and subsequently investigate the optimal tradeoff between efficiency and resilience. The main contributions of this paper can be summarized as follows.

- This study offers a novel ANPR sensing network solution in terms of systemic redundancy optimization, which provides new evidence for better understanding this spatial optimization problem.
- We develop a two-stage GDL model and introduce graph pooling and convolutional networks in the model construction process, thereby enriching the available toolbox for scholars and practitioners. To the best of our knowledge, this is the first study to apply graph pooling with respect to the systemic optimization of ANPR sensing networks.
- This paper provides insights for balancing efficiency and resilience when managing traffic monitoring sensors; these insights can be generalized and applied to many other networks, such as sensor networks, if given the appropriate context.

II. Literature review

A. GDL in transportation

Convolutional neural networks (CNNs) successfully leverage the properties of image-like data on Euclidean domains [9]. Many studies have been conducted to apply CNNs to capture the spatial features of traffic networks [10], [11]. CNN-based models are suitable for solving the problems that involve modeling the Euclidean correlations among different regions [12], [13]. These studies converted traffic networks to regular grids because the CNNs can only process Euclidean-structured data. Nevertheless, time series on road networks are continuous sequences distributed over a topology graph, which are typical representatives of non-Euclidean-structured data. To fill this gap, attempts have been made to extend convolution and pooling layers in the non-Euclidean domain [14], [15]. These graph convolution and pooling layers are also known as graph neural network (GNN) layers.

Graph convolution layers have been widely used to process graph data in the field of transportation. In [16], a temporal graph convolutional network (TGCN) combining recurrent neural networks (RNNs) [17] and graph convolutional networks (GCNs) [18] was proposed for traffic prediction. [19] considered the traffic flow as a diffusion process on a directed graph and developed a diffusion-based convolutional RNN (DCRNN) for traffic forecasting. [20] developed a spatiotemporal GCN (STGCN) that uses GCNs and gated CNNs for capturing spatial and temporal features, respectively. More recently, several studies introduced attention mechanisms to learn data-dependent graphs. In [21], an attention-based spatialtemporal GCN (ASTGCN) was presented, in which an attention mechanism was employed to extract the dynamic spatial-temporal correlations in traffic data. [22] proposed an attention-based spatiotemporal graph attention network (ASTGAT) for segment-level traffic speed prediction. The authors adopted multiheaded graph attention networks (GATs) [23] and attention-based RNNs to learn spatial and temporal dependencies, respectively.

Fewer studies have considered graph pooling than for graph convolution. Earlier studies utilized either graph topologies [24] or node features [25] to generate a reduced graph representations. Recently, several hierarchical graph pooling methods have been proposed [26], [27]. These methods allow graph convolution layers to attain scaled-down graphs after performing pooling in an end-to-end fashion. However, the differentiable pooling (DiffPool) method [27] has a quadratic storage complexity, and its number of its parameters is dependent on the number of nodes. Although the gPool [14] and top-K pooling [26] methods possess decreased the model complexity, neither method considers graph topology. [28] proposed a self-attention graph pooling (SAGPool) method to address these issues. By employing a graph convolution layer to compute self-attention scores, the result of SAGPool is based on both node features and graph topology. The only trainable parameter of SAGPool is the weight matrix of the employed graph convolution layer. Currently, graph pooling layers are mainly used for learning coarse graph embeddings to facilitate graph classification [29]. Few attempts have been made to introduce graph pooling techniques in the field of transportation research.

B. System optimization

This section summarizes recent works in the field of system optimization. In [30], the deployment problem of edge servers (ESs) deployment problem was optimized by a many-objective evolutionary algorithm. A clustering algorithm was adopted to initialize the populations of the many-objective evolutionary algorithm, which was then used to optimize the proposed many-objective deployment model to achieve a tradeoff between conflicting objectives. A many-objective optimization model was built by [31] to address the multidepot and heterogeneous-vehicle capacitated arc routing problem (CARP). Then, a memetic algorithm was proposed on the basis of TwoArch2 (MATA) to optimize the constructed model. In [32], the authors developed a service pricing-based two-stage incentive algorithm (STIA) to tackle the problems regarding the insufficiency of participating nodes and their low willingness to participate. The study by Mou et al. [33] presented the energy-efficient distributed permutation flow-shop inverse scheduling problem (EEDPFISP) to minimize the processing time adjustment and energy consumption. An effective collaborative algorithm with hybrid initialization approaches and cooperative search operators were designed to solve the EEDPFISP model. To address the redeployment issue of pavement distress detection models, the authors in [34] proposed a CNN-based cross-scene TL pipeline, which combines the advantages of model transfer and data transfer. In [35], two novel and reliable static random-access memory (SRAM) cells were proposed with enhanced self-recoverability from single-node upsets (SNUs) and double-node upsets (DNUs). [36] presented a novel multicriteria group decision making (MCGDM) approach based on intuitionistic fuzzy sets (IFSs) and the VIKOR method for assessing the capacity of COVID-19 medical waste (CWM) recycling channels.

C. ANPR systems and ANPR data

ANPR or automatic license plate recognition (ALPR), leverages image processing technologies to capture images of vehicles and extract the numbers from the plates in these images, and then translates the information into a machine-readable format [37], [38]. Many models and algorithms have been developed to improve the recognition accuracy and enhance the overall robustness of systems against the complex environments, such as those with dirty and blurred plates [39], poor lighting [40], low-resolution images [41], and duplicated plates [42]. The multidimensional information provided by ANPR data can insightfully be used to reconstruct vehicle movements and, thus, to unravel vehicle mobility patterns [43]. ANPR data can also be leveraged to estimate origin-destination (OD) characteristics [44]. For instance, by extracting flow information from ANPR data, [45] devised a 3D CNN approach to reveal high-dimensional correlations between the local vehicle patterns and OD flows. In addition, some scholars have used ANPR data as benchmarks to comparatively assess the performance of other data collection methods in terms of travel time estimation [46] and combined them with other techniques, such as clustering methods, to identify commuting patterns [47]. In general, ANPR data are considered very useful and informative in transportation studies and have been widely applied in many research topics.

Based on the aforementioned literature survey, the main research gaps can be summarized as follows. (1) It is necessary to further explore and develop effective models for optimizing systemic redundancy in fixed ANPR sensor arrays; in particular, graph learning has not been applied in such topics. (2) The resilience of the city-level ANPR sensor networks is of importance for urban management, such as vehicle mobility monitoring and security surveillance by the police force. However, this topic is relatively underdeveloped in the literature. Most importantly, (3) studies on balancing efficiency and resilience in urban transportation infrastructure could potentially provide meaningful insights for sustainable and resource-friendly management during planning and design. Nevertheless, this topic is also scarce in the literature and thus needs to be further explored, especially for ANPR traffic sensing networks.

III. Methodology

A. Problem formulation

In the proposed graph learning framework, we model a road network with N sensors as an undirected graph, $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{A})$, where $\boldsymbol{V}$ is a set of vertices corresponding to observations from N sensors, and the weighted adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{N}}$ depicts the connectivity between the vertices. The adjacency matrix $\boldsymbol{A}$ is computed based on the distances among the sensors in the road network:

$$A_{ij} = \begin{cases} \exp\left(-\dfrac{\text{dist}(s_i, s_j)}{\sigma^2}\right), & \text{if} \exp\left(-\dfrac{\text{dist}(s_i, s_j)}{\sigma^2}\right) \geq \varepsilon \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where A_{ij} represents the edge weight between sensor s_i and sensor s_j, which is decided by their Euclidean spatial distance $\text{dist}(s_i, s_j)$, and σ^2 and ε are the user-controlled parameters that control the distribution and sparsity of the matrix. σ^2 and ε are set as 10 and 0.5 respectively, which are the same as those in [20], [48].

This study focuses on optimizing a city-scale traffic sensing network, which involves two subproblems: 1) selecting the most representative K sensors from the original N ($K < N$) sensors, and 2) estimating the most likely traffic measurements at all N sensors based on the observations from the selected K sensors. These two subproblems can be formulated by

$$V_T \xrightarrow{f_{slc}(\cdot)} \text{idx} \tag{2}$$

$$V_t' \xrightarrow{f_{est}(\cdot)} \hat{V}_t \tag{3}$$

where $V_T \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{T}}$ denotes the observations from $\boldsymbol{N}$ sensors and the past T time steps; idx represents the indices of the selected K sensors; $V_t \in \mathbb{R}^{\boldsymbol{K}\times\boldsymbol{1}}$ denotes the observations obtained from the selected $\boldsymbol{K}$ sensors at time step t; and $\hat{V}_t \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{1}}$ represents the estimated traffic measurements at all $\boldsymbol{N}$ sensors at time step t. The proposed framework aims to learn a selection function $f_{slc}(.)$ and an estimation function $f_{est}(.)$.

B. Sensor selection model

To identify the optimal subset of sensors, we propose a SAGPool-based sensor selection model. SAGPool considers both node features and graph topology and learns the pooling procedure in a self-attention manner. Attention mechanisms allow us to focus more on important features and less on unimportant features. Self-attention, also known as intra-attention, is a particular attention mechanism that allows the input features to be the criteria for the attention module itself [49]. As illustrated in Fig. 1, the SAGPool utilizes a GCN to compute the self-attention score $\boldsymbol{Z} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{1}}$:

$$Z = D^{-1/2} A^I D^{-1/2} X W^{pool} \tag{4}$$

$$A^I = A + I \tag{5}$$

$$D_{ii} = \sum_j A^I_{ij} \tag{6}$$

where A^I is the adjacency matrix with added self-loops; I is the identity matrix; D is the degree matrix of A^I; $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{F}}$ denotes the features of the graph with N nodes and F dimensional features; and W^{pool} is a layer-specific trainable weight matrix.

Following this, the top K nodes are selected based on the obtained self-attention score Z:

$$\text{idx} = \text{top-rank}(Z, K) \tag{7}$$

$$K = rN \tag{8}$$

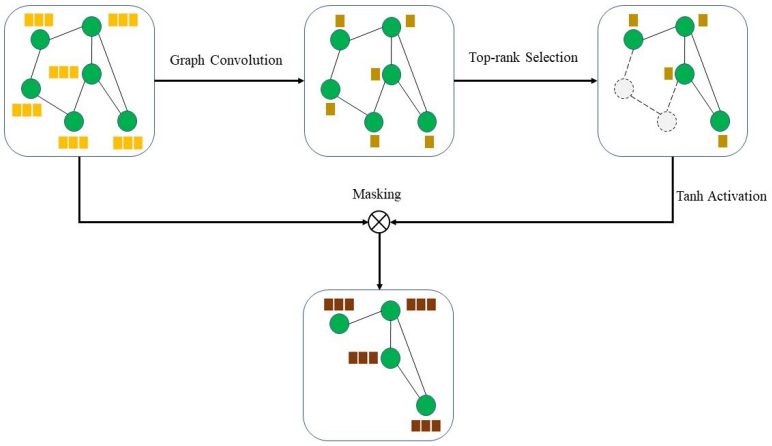


Fig. 1: Illustration of the SAGPool method.

$$Z^{mask} = Z^{\text{idx}} \tag{9}$$

where the pooling ratio $r \in (0, 1]$ is a hyperparameter that determines the number of nodes to keep; $top-rank()$ is a function that returns the indices of the top K values; idx is an indexing operation; and Z^{mask} is the feature attention mask.

The reduced feature matrix $\boldsymbol{X^{out}} \in \mathbb{R}^{\boldsymbol{K}\times\boldsymbol{F}}$ is obtained via the operation denoted as $\boldsymbol{masking}$ in Fig. 1:

$$X^{out} = X^{\text{idx},:} \odot \tanh(Z^{mask}) \tag{10}$$

where $X^{\text{idx},:}$ is the rowwise (i.e. nodewise) indexed feature matrix; $tanh()$ is the $tanh$ activation function; and $\odot$ is the broadcasted elementwise product.

The architecture of the proposed node selection model is shown in Fig. 2. The model input is a single graph defined by a node feature matrix X^{slc} of shape (N, T) and an adjacency matrix A of shape (N, N), where T is the length of the training time series. This single graph is passed to a SAGPool layer:

$$h^{pool}, \text{idx} = \text{SAGPool}(X^{slc}, A) \tag{11}$$

where $\boldsymbol{h^{pool}} \in \mathbb{R}^{\boldsymbol{K}\times\boldsymbol{T}}$ is the output of the SAGPool layer; and idx denotes the indices of the K nodes selected in the SAGPool process.

We design an unpooling layer to perform the inverse operation of the SAGPool layer and restore the graph to its original structure. The layerwise propagation rule of the unpooling layer can be described as:

$$h^{unpool} = \text{distribute}(\text{zero}, h^{pool}, \text{idx}) \tag{12}$$

where $\boldsymbol{h^{unpool}} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{T}}$ is the output of the unpooling layer; zero $\in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{T}}$ is a matrix in which all of the entries are zero; and $distribute()$ is a function that distributes the row vectors in h^{pool} into a zero matrix according to their corresponding indices stored in idx. In h^{unpool}, the row vectors with indices in idx are updated by the row vectors in h^{pool}, while the other row vectors remain zero.

Then, the unpooled feature matrix h^{unpool} is cut into T slices of shape $(N,1)$, $\{h_1^{unpool}, ..., h_t^{unpool}, ..., h_T^{unpool}\}$, which are passed to a fully connected (FC) layer with N units:

$$h_t^{fc} = \text{FC}(h_t^{unpool}) \tag{13}$$

where h_t^{fc} is the t^{th} output of the FC layer. Note that we apply transpose operations to the inputs and outputs of the FC layer to meet the input shape requirement of the layer and restore the shape of the feature matrix, respectively.

Finally, we concatenate all the outputs of the FC layer and obtain the output of the node selection model, $\boldsymbol{y^{slc}} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{T}}$. Note that y^{slc} is only used for model training. The main purpose of the selection model is to locate a subset of nodes (i.e., idx from the SAGPool layer) that minimizes the difference between y^{slc} and the real observations.

C. Estimation model

The estimation model aims to estimate traffic measurements at all N sensors given the information from the selected K sensors. Fig. 3 illustrates the proposed estimation model, which is a simple 2-layer neural network. At each time step t, an input feature matrix $\boldsymbol{X_t^{est}} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{1}}$ is built, in which the row vectors without indices in idx are filled with zeros. The input graph is processed by a GCN layer with 1 channel:

$$h_t^{gcn} = \text{GCN}(X_t^{est}, A) \tag{14}$$

where $\boldsymbol{h_t^{gcn}} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{1}}$ is the output of the GCN layer.

Then, an FC layer is added to generate the estimated traffic measurements at time step t:

$$y_t^{est} = \text{FC}(h_t^{est}) \tag{15}$$

where $\boldsymbol{y_t^{est}} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{1}}$ is the output of the FC layer. It is worth noting that we keep the structure of the estimation model as parameter-economical as possible. One can always introduce more complex network structures to further improve model performance.

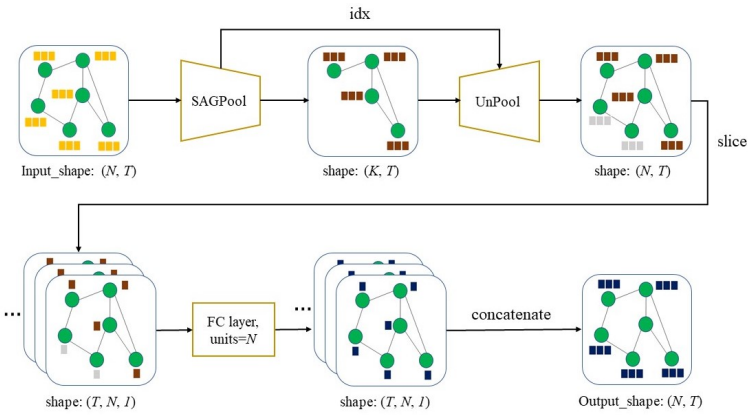


Fig. 2: Architecture of the nodel selection model. The input is passed to a SAGPool layer, and an unpooling layer performs the inverse operation; then, the unpooled features are fed to an FC layer; finally, all the FC outputs are concatenated.

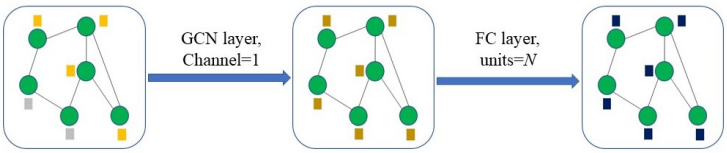


Fig. 3: Architecture of the estimation model. The input graph is processed by a GCN layer with 1 channel, which is followed by an FC layer that is used to generate the estimation output.

D. Resilience assessment metric

For networked systems, the overall resilience can be evaluated using the systemic performance loss, known as the 'resilience curve'. The 'failure-recovery' process illustrated in the Fig. 4 is a typical resilience curve [50]. In this study, the key performance indicator (KPI) on the y-axis is selected as the overall observability of the ANPR system; it depicts how many vehicles can be successfully captured by the total ANPR sensors. Thus, the higher the observability is, the better the performance of the ANPR system.

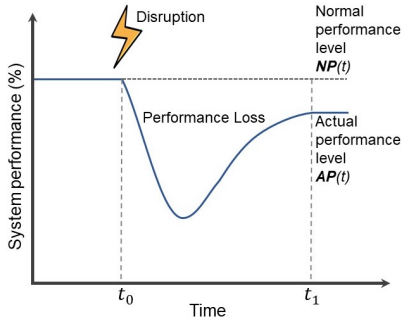


Fig. 4: An illustrative example of the resilience curve paradigm. In this case, the post-event performance level is lower than the pre-event level, which demonstrates an insufficient post-event recovery. Modified based on [51]

.

We adopt the well-applied resilience quantification metric proposed by [51] in this paper. The reason for doing so is twofold: (1) it has an intuitive value range, $[0, 1]$, which facilitates comparative analyses in the later stage, and (2) it exhibits excellent effectiveness and simplicity, as the metric features wide applications in various infrastructure systems and use cases. The metric is defined as follows.

$$RI = \frac{\int_{t_0}^{t1} AP(t)dt}{\int_{t_0}^{t1} NP(t)dt} \tag{16}$$

where t_1 is the total time period of one disturbance event in the system, $AP(t)$ is the actual level of performance at time t considering the impact from the disturbance, and $NP(t)$ is the normal or expected level of performance at time t, which by default is normally at 100%, indicating a fully functioning system.

IV. Specifications and Data

A. Study area

The selected study area is Cambridge (the county town of Cambridgeshire, UK), which is located approximately 80km north of London. The datasets contain detailed trip chain information derived from the ANPR sensors which can be found at Cambridgeshire Insight Open Data. There are 96 camera sensors deployed within the study area with labeled IDs ranging from 1 to 96. However, due to severe missing entries and discontinuous recordings, 90 sensors are available for this study. Fig. 5 shows the geographic layout of the study area and the locations of the ANPR sensors. Vehicle data were collected from June 11th to 17th, 2017 (7 days from Sunday to the following Saturday). All camera sensors recorded continuously from 00:00:00 to 23:59:00 throughout the day to capture all recognizable vehicles and their trips

based on the temporal trajectory stamps within the area. Vehicle IDs and sensor IDs were also specifically recorded.

B. Experimental setup

In this study, the 5-weekday data obtained from 12/06/2017 to 16/06/2017 are collected and divided into two subsets. The data from the first three days are used as the training dataset, and the rest of the data are used as the testing dataset. The training and testing datasets are separately normalized to [-1,1] using Z Score normalization. In the testing dataset, only the observations from the selected nodes are kept. After performing estimation, the estimated values are denormalized and then used for evaluation.

The selection model is trained using the Adam optimizer [53] to minimize the MAE. The learning rate is set to 1×10^{-3}. Note that the input of the selection model is a single graph including all the training time series; therefore, no validation dataset is used. The estimation model is trained using the Adam optimizer (learning rate=1×10^{-3}) to minimize the RMSE. 30% of the training dataset is selected as the validation dataset. Early stopping is used on the validation dataset, and the batch size is set to 32.

V. Results and discussion

The analysis workflow first completes the redundancy optimization process and assesses the model performance using error indicators. After that, we quantify the systemic resilience for each optimization outcome and then locate the tradeoff point between these two. Note that the systemic resilience is calculated based on the redundancy optimization outcomes, forming a constrained failure simulation in resilience analysis. In other words, which sensors will be removed from the network (or masked in this case) to evaluate the resilience are already determined by the optimization process beforehand.

A. Optimization outcomes

To review the network redundancy by applying the developed two-stage GDL model, sensors are gradually masked in 10% increments from 100% (i.e., all 90 unmasked sensors in total) to only 10% (i.e., only 9 unmasked sensors). Fig. 6 shows the geographical changes of the sensor network from 100% to 70% mask rates in the optimization results. As the number of masked sensors increases, it is clear that those redundant sensors concentrate on the periphery of the city. At 90% coverage (i.e., 10% of the sensors are masked), the two most distant sensors at the northwest corner and several sensors on the west side of the M11 motorway are all masked. By increasing the number of masked sensors to 70%, most of the sensors along M11 are masked (i.e., identified as system redundancies by the model), which indicates that the main traffic flow characteristics at those sensors along the motorway can be mostly represented by just several sensors, which is intuitive for conventionally interpreting the flow on the motorway without signal control having more consistent patterns. Fig. 7 shows the results obtained for coverage rates from 60% to 10%. Here, the layout of those masked sensors gradually percolates from the periphery to the central area. In the case with 10% coverage, only nine sensors remain in the network; this is apparently not an efficient outcome for learning the full knowledge of the vehicle movements in the city.

B. Estimation errors

Fig. 8 shows the selection and estimation errors induced by the developed model under two error indicators as the number of unmasked sensors drops (i.e., the MAE and RMSE); a clear increasing trend can be observed. The errors derived from the selection stage are higher than those from the estimation stage. Before 60%, the selection error increases more rapidly than the estimation error. Taking the estimation error as a reference, the error trend starts to become steeper after 50% (Fig. 8(a)). At 40%, the error reaches approximately 14 according to the MAE indication, meaning that approximately 14 vehicles would be over/underestimated from each sensor recording on average, and this number reaches its highest level of approximately 18 at 10% coverage, indicating that the model performance is unreliable below 40% of the unmasked rate. A similar tendency can also be observed using the RMSE as the indicator.

To verify the accuracy of the model estimation process before it becomes unreliable at 40%, we further analyze the discrepancy between the ground-truth data and the estimated results for all sensors during two typical weekday peak-hour periods under the 50% mask rate. Fig. 9 illustrates the former by comparing the model estimations with the ground-truth vehicle counts provided by all 90 sensors at 8am (morning peak hour, subplot (a)) and 5 pm (evening peak hour, subplot (b)) on 15/06/2017 (Thursday). The first observation is that the overall pattern for these 90 sensors is quite similar during both morning and evening peak hours, while some sensors have larger fluctuations in terms of the number of captured vehicles, such as ID 35 and 65. In terms of the estimation accuracy, both subplots demonstrate fairly good estimation accuracy (with a discrepancy of approximately 9 vehicles per sensor on average, which is comparatively acceptable). However, we can also identify larger discrepancies at certain sensors, such as ID 8 and 41. A similar pattern can also be observed for the evening peak hour.

C. Sensitivity and model comparison

To test the model sensitivity and further validate the model performance, we compare the developed model with five different combinations of benchmarks, namely, two models for the selection stage and three models for the estimation stage. This combination yields six types of models in total, which can be elaborated as follows.

1) **Model 0**: This is our developed model, which has the SAGPool submodel for the selection and the one-channel GCN for the estimation. Model 0 is compared with the other five models to test its sensitivity and superiority.
2) **Model 1**: This model has the SAGPool submodel for the selection process but has an eight-channel GCN

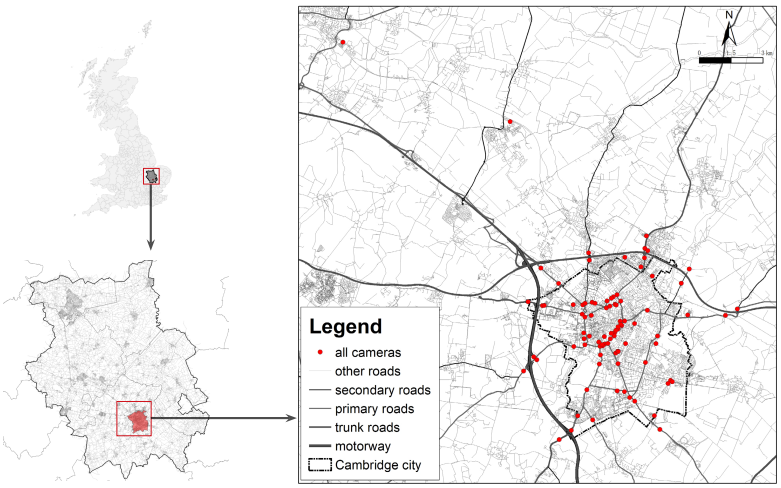


Fig. 5: The study area and overview of the ANPR sensors (The base road map can be downloaded from OpenStreetMap. The location data of the ANPR sensors can be downloaded from [52]).

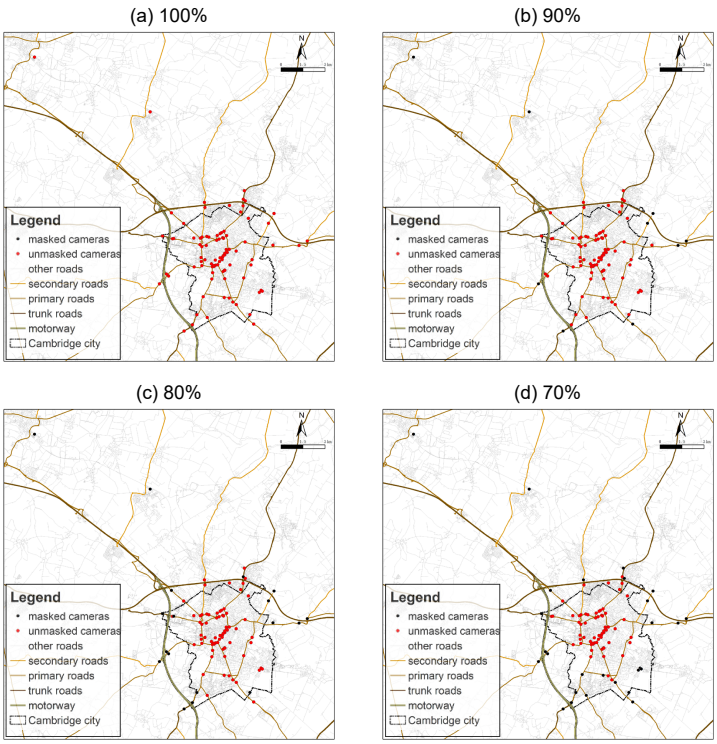


Fig. 6: Changes in geographical layout with descending mask rates from 100% to 70%.

before the final one-channel GCN for the estimation step to view the performance achieved under a variant model complexity when extracting underlying feature representations.

3) **Model 2**: This model still has the SAGPool submodel for selection but contains a diffusion graph convolutional network (DGCN) layer for estimation purposes. The implementation method of the DGCN is directly adopted from [54].
4) **Model 3**: This model changes the selection component from SAGPool to a TopKPool model. The estimation model remains the same as that of **Model 0**. The implementation of the TopKPool model is based on [14].
5) **Model 4**: This model combines TopKPool with the selection model from **Model 1**.
6) **Model 5**: This model combines TopKPool with the selection model from **Model 2**.

Table I presents the error results yielded by the six models

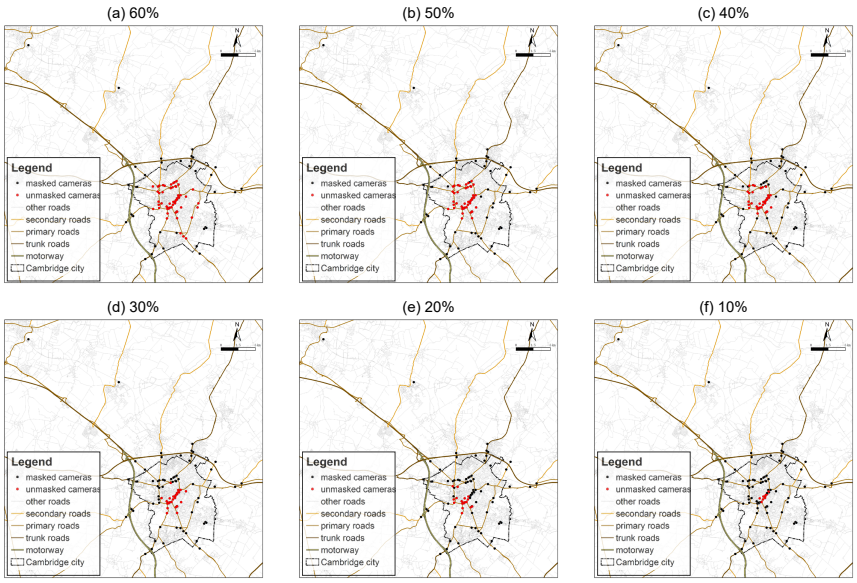


Fig. 7: Further testings results regarding the optimization effect with descending mask rates from 60% to just 10%.

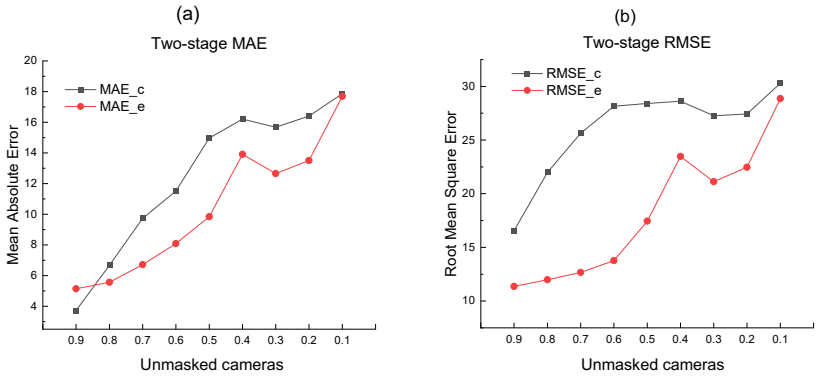


Fig. 8: Comparison of the estimation errors between the model estimations and the groud-truth vehicle counts. Note: "c" represents the accuracy of the selection results, and "e" indicates the accuracy of the estimation results.

on the testing set for both the selection and estimation stages under a 50% mask rate. It is clear that Model 0 outperforms the other five comparative models in both the selection and estimation stages. Model 2 obtains the second-best estimation results, with an MAE of 9.95 and an RMSE 17.71. In contrast, Model 4 is associated with the highest errors of all (Model 1 also has close values), with a selection error of 16.54 and an estimation error of 11.12 in terms of the MAE. The comparison between Models 0 and 1 indicates that the developed Model 0 has higher efficiency, as adding more GCN channels in the estimation stage does not improve the overall accuracy but instead harms it.

Fig. 10 shows the performance achieved by our model with different hyperparameters. Note that the default hyperparameters of the proposed model are as follows: number of GCN channels = 1, SAGPool activation = tanh and kernel initializer = glorot uniform. A 50% mask rate is utilized for the benchmark test. It is observed that the model with

TABLE I: Errors from all comparative models

	Selection stage		Estimation stage	
	MAE	RMSE	MAE	RMSE
Model 0	**14.97**	**28.41**	**9.84**	**17.45**
Model 1	14.97	28.41	11.08	20.76
Model 2	14.97	28.41	9.95	17.71
Model 3	16.54	28.86	10.02	18.10
Model 4	16.54	28.86	11.12	20.87
Model 5	16.54	28.86	10.07	18.87

the default settings yields the best performance. One possible reason behind this is that with the increase in the number of channels, the model may learn irrelevant features (even noise and random fluctuations) from the training data, which decreases the generalization of the model (i.e., the model's ability to react to new data), and results in the model overfitting. Simply increasing the number of GCN channels worsens the model performance. The SAGPool module with tanh activation

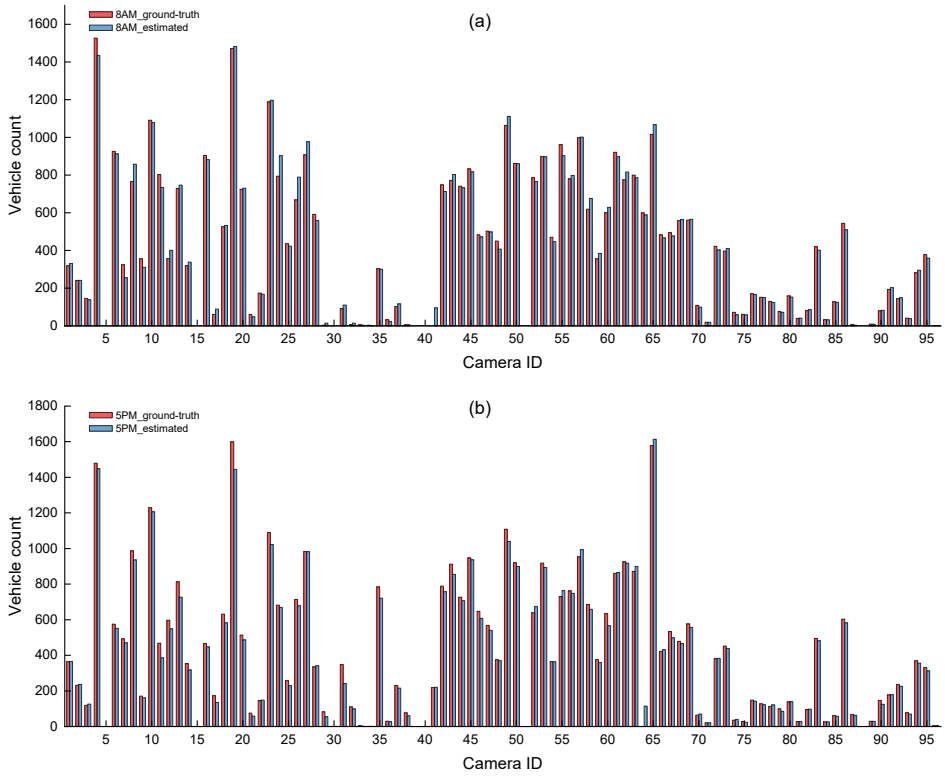


Fig. 9: Comparison between the ground-truth counts and the estimated counts for all 90 sensors at the 50% of masked rate on 15/06/2017 during the morning and evening peak hours.

fits our case better than the SAGPool module with the sigmoid function. Implementing kernel initializers, which are used to statistically initialize the weights in the model, produces better estimation results.

D. Resilience analysis and tradeoff identification

Masking sensors from the network can inevitably cause an overall observability loss, which further deteriorates the systemic resilience of the network in terms of its monitoring capability. As mentioned above at the beginning of the Section V, the resilience analyses here are constrained simulations with predetermined masked sensors based on the optimization outcome in the previous stage. Given this prerequisite, we can only apply simulations to obtain the removal sequence; i.e., only two types of removal scenarios are applicable for the predetermined removed sensors in each simulation iteration: being preferentially ordered or randomly selected (note: thus, other types of removal, such as localized removal due to natural hazards, are not applicable in this study). For example, under a given mask rate of 90%, which 10% of sensors will be removed are already determined by the optimization outcome, and the resilience analysis simulates these two removal sequences for those 10% masked sensors.

Fig. 11 shows the simulation results obtained under each mask rate, and two prominent features can be observed. First, the ordered removal process demonstrates a smoother curve than random removal under all mask rates, as the former removes sensors by following the individual observability values in descending order. The curves of the random removals are fuzzier, forming a more regular triangle shape in each curve. Second, in terms of the overall tendencies of all curves, the difference between these two types of removals is trivial, and this is confirmed by the calculated resilience indices: subplots (b) and (d) demonstrate two rather similar decreasing lines for the resilience indices. A numerical check also confirms that these two scenarios yield basically the same results, with minor discrepancies after two decimal places. Regarding the sensors to be removed being predetermined by the redundancy optimization results, the removal sequences of the masked sensors (either ordered or random) do not significantly alter the total performance loss.

Given that the main purpose of the proposed model is to reduce the number of redundant sensors in the array to save costs in the total design and planning budget, the target of the optimization process here is to reduce the sensor quantity as much as possible (i.e., to maximize the cost savings) while maintaining a good level of systemic resilience. After obtaining the outcomes of both optimization and resilience tests, the dilemma problem here becomes a simple linear programming issue, which can be solved graphically. Fig. 12 depicts the solution profiles of this linear programming problem with both the MAE and RMSE indicators. The optimal location is at the intersection, where the unmasked rate is 0.47 (meaning that 53% of the total sensors can be masked as redundancies) and the resilience is 0.7 (i.e., 70% of resilience capacity) with an MAE of 11.18. This result is of great interest, as it indicates that by applying the proposed GDL model, this particular ANPR sensor network can be optimized by remov-

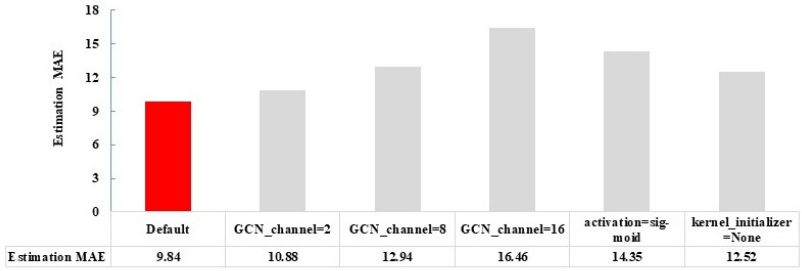


Fig. 10: Hyperparameter test results obtained under a 50% mask rate. Default hyperparameters: number of GCN channels = 1, SAGPool activation function = tanh and kernel initializer = glorot uniform.

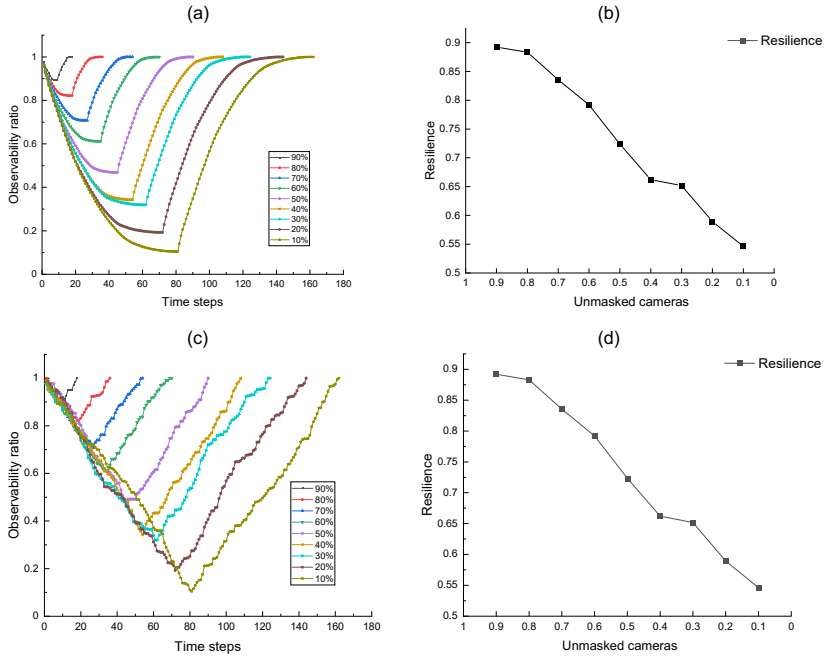


Fig. 11: Resilience assessment under deliberate (a, b) and random (c, d) ordered scenarios. Note that subplots (a) and (c) are resilience curves while (b) and (d) are calculated resilience indices.

ing approximately 47 sensors in total while still maintaining a relatively high level of resilience and a moderate estimation error level; this can potentially save significant costs for this city-scale ANPR scheme in Cambridge. To demonstrate, as claimed by [55], the annual cost of London's 1666 ANPR sensors is approximately 5.9m GBP, which leads to a rough estimation of 3541 GBP per sensor per year for the ANPR management budget. In this case, removing approximately 47 sensors represents a total budget reduction of approximately 166k GBP for Cambridge's ANPR scheme.

VI. Conclusions

This paper investigates the problem of optimizing systemic redundancy while maintaining the resilience in sensor networks. Taking the city-scale ANPR system in Cambridge, UK, as an illustrative example, we demonstrate that this problem can be effectively approached with a graphical solution. The main conclusions can be summarized as follows.

- We develop a novel two-stage GDL model that takes both node features and graph topology into account to achieve an optimal subset of nodes. The estimation results are acceptable at both the systemic and individual levels, indicating that the developed model is an effective tool for optimizing the systemic redundancy in this case study.
- In the model comparison and sensitivity analysis, our model outperforms all five comparative models with an MAE of only 9 vehicles during the estimation stage; it demonstrates excellent performance in terms of feature reconstruction in this geographical optimization problem.
- We find the tradeoff point at which only 47% of the total sensors are used while still maintaining the overall systemic resilience at the 70% level with an acceptable estimation error, indicating a great reduction in systemic redundancy while still maintaining fairly strong robustness against the risk of performance loss, and this outcome could potentially yield significant cost savings for ANPR schemes.

We also acknowledge two limitations of this study. First, due to the data availability issue, we only use one week of ANPR data. Second, this study only used one ANPR case

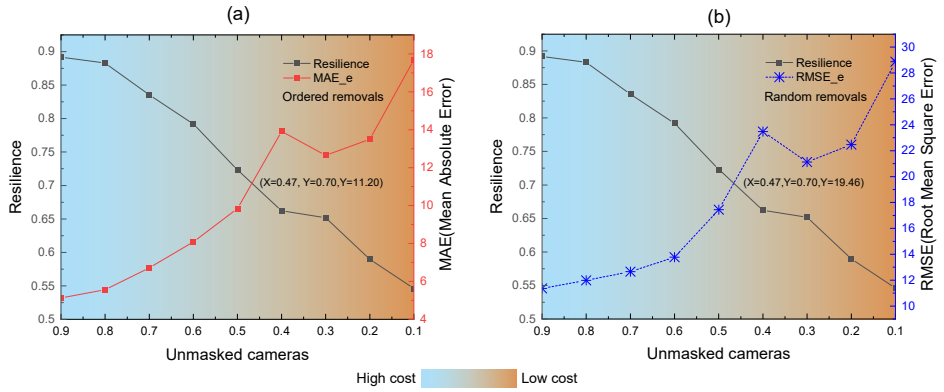


Fig. 12: Efficiency vs. resilience under ordered and random removals.

study. For the case study, other cities should be tested. Our future work will focus on addressing these two limitations.

REFERENCES

[1] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustainable Cities and Society*, vol. 38, pp. 697–713, 2018.

[2] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60–70, 2016.

[3] S. Sarkar, M. W. Totaro, and K. Elgazzar, "Intelligent drone-based surveillance: application to parking lot monitoring and detection," in *Unmanned Systems Technology XXI*, vol. 11021. International Society for Optics and Photonics, 2019, p. 1102104.

[4] N. J. Bertola, A. Costa, and I. F. Smith, "Strategy to validate sensor-placement methodologies in the context of sparse measurement in complex urban systems," *IEEE Sensors Journal*, vol. 20, no. 10, pp. 5501–5509, 2020.

[5] M. Parsafard and X. Li, "Sensor location design for interdicting mobile travelers with probabilistic space-time trajectories," *Transportation Research Part C: Emerging Technologies*, vol. 132, p. 103420, 2021.

[6] J. Tang, L. Wan, T. Nochta, J. Schooling, and T. Yang, "Exploring resilient observability in traffic-monitoring sensor networks: A study of spatial–temporal vehicle patterns," *ISPRS International Journal of Geo-Information*, vol. 9, no. 4, p. 247, 2020.

[7] X. Xu, A. Chen, S. Jansuwan, C. Yang, and S. Ryu, "Transportation network redundancy: Complementary measures and computational methods," *Transportation Research Part B: Methodological*, vol. 114, pp. 68–85, 2018.

[8] X. Zhang and H. Li, "Urban resilience and urban sustainability: What we know and what do not know?" *Cities*, vol. 72, pp. 141–148, 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[11] J. Ke, H. Zheng, H. Yang, and X. M. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017.

[12] R. Jiang, Z. Cai, Z. Wang, C. Yang, Z. Fan, X. Song, K. Tsubouchi, and R. Shibasaki, "Vluc: An empirical benchmark for video-like urban computing on citywide crowd and traffic prediction," *arXiv preprint arXiv:1911.06982*, 2019.

[13] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 5668–5675.

[14] H. Gao and S. Ji, "Graph u-nets," in *international conference on machine learning*. PMLR, 2019, pp. 2083–2092.

[15] F. M. Bianchi, D. Grattarola, and C. Alippi, "Spectral clustering with graph neural networks for graph pooling," in *International Conference on Machine Learning*. PMLR, 2020, pp. 874–883.

[16] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2019.

[17] J. Schmidhuber and S. Hochreiter, "Long short-term memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[19] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.

[20] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[21] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 922–929.

[22] D. Li and J. Lasenby, "Spatiotemporal attention-based graph convolution network for segment-level traffic prediction," *IEEE Transactions on Intelligent Transportation Systems (Accepted)*, 2021.

[23] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[24] S. Rhee, S. Seo, and S. Kim, "Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification," *arXiv preprint arXiv:1711.05859*, 2017.

[25] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, "An end-to-end deep learning architecture for graph classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[26] C. Cangea, P. Veličković, N. Jovanović, T. Kipf, and P. Liò, "Towards sparse hierarchical graph classifiers," *arXiv preprint arXiv:1811.01287*, 2018.

[27] R. Ying, J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *arXiv preprint arXiv:1806.08804*, 2018.

[28] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *International Conference on Machine Learning*. PMLR, 2019, pp. 3734–3743.

[29] J. Qin, L. Liu, H. Shen, and D. Hu, "Uniform pooling for graph networks," *Applied Sciences*, vol. 10, no. 18, p. 6287, 2020.

[30] B. Cao, S. Fan, J. Zhao, S. Tian, Z. Zheng, Y. Yan, and P. Yang, "Large-scale many-objective deployment optimization of edge servers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3841–3849, 2021.

[31] B. Cao, W. Zhang, X. Wang, J. Zhao, Y. Gu, and Y. Zhang, "A memetic algorithm based on two_arch2 for multi-depot heterogeneous-vehicle capacitated arc routing problem," *Swarm and evolutionary computation*, vol. 63, p. 100864, 2021.

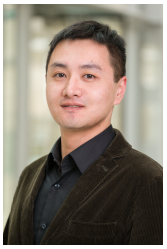
[32] X. Zenggang, L. Xiang, Z. Xueming, Z. Sanyuan, X. Fang, Z. Xiaochao, W. Yunyun, and Z. Mingyang, "A service pricing-based two-stage incentive algorithm for socially aware networks," *Journal of Signal Processing Systems*, pp. 1–16, 2022.

[33] J. Mou, P. Duan, L. Gao, X. Liu, and J. Li, "An effective hybrid collaborative algorithm for energy-efficient distributed permutation flowshop inverse scheduling," *Future Generation Computer Systems*, vol. 128, pp. 521–537, 2022.
[34] Y. Li, P. Che, C. Liu, D. Wu, and Y. Du, "Cross-scene pavement distress detection by a novel transfer learning framework," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 11, pp. 1398–1415, 2021.
[35] A. Yan, Y. Chen, Y. Hu, J. Zhou, T. Ni, J. Cui, P. Girard, and X. Wen, "Novel speed-and-power-optimized sram cell designs with enhanced self-recoverability from single-and double-node upsets," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 12, pp. 4684–4695, 2020.
[36] S. Liu, J. Zhang, B. Niu, L. Liu, and X. He, "A novel hybrid multi-criteria group decision-making approach with intuitionistic fuzzy sets to design reverse supply chains for covid-19 medical waste recycling channels," *Computers & Industrial Engineering*, p. 108228, 2022.
[37] S. Singh, "Optical character recognition techniques: a survey," *Journal of emerging Trends in Computing and information Sciences*, vol. 4, no. 6, pp. 545–550, 2013.
[38] A. Puranic, K. Deepak, and V. Umadevi, "Vehicle number plate recognition system: a literature review and implementation using template matching," *International Journal of Computer Applications*, vol. 134, no. 1, pp. 12–16, 2016.
[39] V.-G. Nguyen and D. L. Nguyen, "Joint image deblurring and binarization for license plate images using deep generative adversarial networks," in *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*. IEEE, 2018, pp. 430–435.
[40] P. Łubkowski and D. Laskowski, "Assessment of quality of identification of data in systems of automatic licence plate recognition," in *International Conference on Transport Systems Telematics*. Springer, 2017, pp. 482–493.
[41] H.-s. Min, S. H. Lee, W. De Neve, and Y. M. Ro, "Improved license plate recognition for low-resolution cctv forensics by integrating sparse representation-based super-resolution," in *International Workshop on Digital Watermarking*. Springer, 2013, pp. 452–462.
[42] A. J. Hoffman, D. J. Geldenhuys, and A. B. Pretorius, "Securing number plates based on digital signatures and rfid," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, 2013, pp. 2051–2057.
[43] A. H. Chow, A. Santacreu, I. Tsapakis, G. Tanasaranond, and T. Cheng, "Empirical assessment of urban traffic congestion," *Journal of advanced transportation*, vol. 48, no. 8, pp. 1000–1016, 2014.
[44] A. Robinson and C. Venter, "Validating traffic models using large-scale automatic number plate recognition (anpr) data," *Journal of the South African Institution of Civil Engineering*, vol. 61, no. 3, pp. 45–57, 2019.
[45] K. Tang, Y. Cao, C. Chen, J. Yao, C. Tan, and J. Sun, "Dynamic origin-destination flow estimation using automatic vehicle identification data: A 3d convolutional neural network approach," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 30–46, 2021.
[46] E. Purson, E. Klein, A. Bacelar, F. Reclus, and B. Levilly, "Simultaneous assessments of innovative traffic data collection technologies for travel times calculation on the east ring road of lyon," *Transportation Research Procedia*, vol. 6, pp. 79–89, 2015.
[47] R. Hong, W. Rao, D. Zhou, C. An, Z. Lu, and J. Xia, "Commuting pattern recognition using a systematic cluster framework," *Sustainability*, vol. 12, no. 5, p. 1764, 2020.
[48] C. Zhang, J. James, and Y. Liu, "Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting," *IEEE Access*, vol. 7, pp. 166 246–166 256, 2019.
[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.
[50] C. W. Zobel and L. Khansa, "Characterizing multi-event disaster resilience," *Computers & Operations Research*, vol. 42, pp. 83–94, 2014.
[51] M. Ouyang and L. Duenas-Osorio, "Time-dependent resilience assessment and improvement of urban infrastructure systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 3, p. 033122, 2012.
[52] Cambridgeshire Insight Open Data, "Greater cambridge anpr data," Cambridgeshire County Council, Tech. Rep., 2020. [Online]. Available: https://data.cambridgeshireinsight.org.uk
[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[54] Y. Wu, D. Zhuang, A. Labbe, and L. Sun, "Inductive graph neural networks for spatiotemporal kriging," *arXiv preprint arXiv:2006.07527*, 2020.
[55] Transport for London, "How much does it cost annually to maintain anpr cameras operated by transport for london?" Greater London Authority, Tech. Rep., 2022. [Online]. Available: https://www.london.gov.uk/questions/2016/3111

Dr. Junqing Tang is an Assistant Professor in Urban system resilience and sustainability research at School of Urban Planning and Design, Peking University. Prior to his faculty position, he was a Postdoc Research Associate at the Centre for Smart Infrastructure and Construction at the University of Cambridge. He obtained his PhD from the Department of Environmental Systems Science at ETH Zurich in 2019, MSc in Transport from Imperial College London and University College London in 2014, and BEng in Architectureal Engineering from Cardiff University in 2013. His research focuses on how resilience and sustainability can be effectively embedded and managed in urban systems.

Shufen Wei is a postgraduate student at School of Urban Planning and Design, Peking University since 2021. She has been supervised by Dr. Junqing Tang and her research interest focuses on investigating infrastructure resilience in cities.

Xiaowei Li works at Xi'an University of Architecture and Technology. He received his Ph.D. from Chang'an University and after that completed his postdoctoral research at Southeast University. His main research interests is focusing on the modeling and analysis of transportation systems using big data and artificial intelligence.



Duo Li (Senior Member, IEEE) received his PhD, M.S. and B.E. Degrees in Civil Engineering (Transportation) from the University of Auckland in 2015, the University of Queensland in 2011, and the Huazhong University of Science and Technology in 2010, respectively. Since 2022, he has been a Senior Lecturer with the Department of Engineering, Nottingham Trent University. Before this, he had several academic and research positions inculding Research Associate at the University of Cambridge, Humboldt Research Fellow at the German Aerospace Center (DLR), and academic positions at the Chang'an University. His Research interests include Intelligent Transport System (ITS) optimization, microscopic and macroscopic transport modelling and simulation, and data-driven traffic analytics and forecasting.