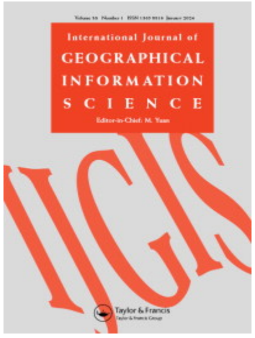

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Estimating urban noise along road network from street view imagery

Jing Huang, Teng Fei, Yuhao Kang, Jun Li, Ziyu Liu & Guofeng Wu

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2024, VOL. 38, NO. 1, 128–155
https://doi.org/10.1080/13658816.2023.2274475

RESEARCH ARTICLE

Estimating urban noise along road network from street view imagery

Jing Huang[a], Teng Fei[a], Yuhao Kang[b], Jun Li[a], Ziyu Liu[a] and Guofeng Wu[c,d]

[a]School of Resource and Environment Sciences, Wuhan University, Wuhan, China; [b]GISense Lab, Department of Geography, University of South Carolina, Columbia, SC, USA; [c]Department of Urban Informatics, Shenzhen University, Shenzhen, China; [d]Guangdong-Hong Kong-Macau Joint Laboratory for Smart Cities, Guangdong, Hong Kong

ABSTRACT

Estimating road traffic noise is essential for examining the quality of sounding environment and mitigating such a non-negligible pollutant in urban areas. However, existing estimated models often have limited applicability to specific traffic conditions, while the required parameters may not be readily available for city-wide collection. This paper proposes a data-driven approach for measuring road-level acoustic information of traffic with street view imagery. Specifically, we utilize portable vehicle-equipped hardware for in-situ noise acquisition and employ a deep learning model ResNet to learn high-level visual features from street view images that are closely associated with road traffic noise. The ResNet captures meaningful patterns from the input data, and the output probability vectors are then fed into a Random-Forest regression algorithm to quantitatively estimate the noise in decibels for different road segments. The MAE and RMSE of the DCNN-RF model are 2.01 and 2.71, respectively. Additionally, we employ a gradient-weighted Class Active Mapping approach to visually interpret our deep learning model and explore the significant elements in streetscapes that contribute to the model's estimations. Our proposed framework facilitates low-cost and fine-scale road traffic noise estimations and sheds light on how auditory information could be inferred from street imagery, which may benefit practices in geography and urban planning.

ARTICLE HISTORY
Received 23 April 2023
Accepted 18 October 2023

KEYWORDS
Road traffic noise; street view imagery; deep learning; build environment; urban planning

1. Introduction

Traffic noise along the road network is a primary source of urban noise pollution, which has adversely affected the urban ecological environment and human health (Khan *et al.* 2018). To control and minimize the risk brought by urban road traffic noise, it is vital to create traffic noise maps. Traffic noise maps present the spatial distribution of noise exposure levels (Murphy and King 2010), and can provide valuable and spatially explicit information to traffic authorities to implement appropriate solutions for noise abatement and

CONTACT Teng Fei feiteng@whu.edu.cn

control, such as paving roads with low-noise surface materials, setting suitable road speed limits, and setting up facilities such as roadside noise barriers and 'No horn' traffic signs at appropriate places. Furthermore, traffic noise maps could help researchers to analyze the potential health risks and negative effects on exposed residents, such as annoyance (Paiva *et al*. 2019), sleep disturbances (Kim *et al*. 2012), and hearing impairments (Kim *et al*. 2012).

In order to obtain the spatial distribution of urban road traffic noise, there are two primary categories of methods: *measurement*-based methods and *modeling*-based methods. The former performs either stationary observation or mobile measurement of noise level at multiple places (Alam *et al*. 2020), while the latter estimates noise based on mathematical modeling of the environment, the sources, and the propagation law of sound (Garg and Maji 2014).

Measurement-based studies have been conducted on noise acquisition (Yilmaz and Hocanli 2006, Mehdi *et al*. 2011), quantification (Doygun and Kuşat Gurun 2008, Li *et al*. 2002), gradation (Banerjee *et al*. 2008, Al-Ghonamy 2009), classification (Barrigón Morillas *et al*. 2005, Mansouri *et al*. 2006, Zambon *et al*. 2016), and source backtracking (Talotte *et al*. 2003, Singh and Davar 2004). In particular, the traffic noises have been studied at different road forms separately, such as at different road grades (Brambilla *et al*. 2020), pavement types (Freitas *et al*. 2012, Son *et al*. 2014), intersections (Okonofua *et al*. 2016), overpasses (Gao *et al*. 2015), and roadworks sites (Tah-Chew and Keung 1991). In addition, a number of key transport facilities have also emerged as research hotspots, such as subway exits (Gershon *et al*. 2006, Neitzel *et al*. 2009, Ghotbi *et al*. 2012), railway stations (Trombetta Zannin and Bunn 2014, Asfiati *et al*. 2020), and airports (Wallis 1997, Cohen *et al*. 2008). Measurement-based methods are generally used for carrying out assessments of noise pollution by calculating traffic noise index or generating noise maps with interpolation techniques (Can *et al*. 2014, Harman *et al*. 2016, Aumond *et al*. 2018). However, such methods face several challenges, including time, cost, and labor-consuming, have a low level of automation, and may lack the spatial and temporal representativeness of the noise samples.

Instead, *modeling*-based methods exhibit a more feasible and efficient way to circumvent these abovementioned limitations. The characteristics of road traffic noise in different locations can be measured using estimated models under different acoustic environmental conditions. Generally, there are two main kinds of modeling-based methods: empirical models (also called statistical models) and machine-learning models. Empirical models, such as FHWA in the USA (Barry and Reagan 1978), ASJ RTN 2008 in Japan (Yamamoto 2010), CoRTN in the UK (Givargis and Mahmoodi 2008), RLS 90 in Germany (Quartieri *et al*. 2009), Son Road in Switzerland (Heutschi 2004), Harmonoise in Europe (Watts 2005), etc., integrate the relationships among a set of traffic-related factors and road conditions and consider the propagation and reduction of noise in air, which brought in regular usage by local and national road authorities in different regions. However, these models follow the assumptions of linear relationships among predictors, which constrain their ability to capture complex nonlinear relationships and interactions.

Compared with the classical statistical models, neural network-based models have nonlinear fitting capabilities, and yield better outcomes (Nedic *et al*. 2014, Bravo-Moncayo *et al*. 2019). In these models, researchers mainly considered traffic volume and velocity as crucial input factors. Several other factors such as longitudinal gradients of the road (Chen *et al.* 2021), roadway temperature (Hamad *et al.* 2017), and honking occurrences (Singh *et al.* 2021) have also been incorporated. However, accurately predicting road traffic noise is a significant challenge due to the complex physical mechanism involved. The multitude of factors related to noise generation and propagation are high-dimensional and of a large scale, which highlights the pressing need to enhance the performance and generalizability of these models. In recent years, deep learning has shown remarkable success in various tasks, emerging as a promising method for conducting traffic-related research. Convolutional neural network (CNN) has been applied to estimate traffic states. For instance, Peppa *et al.* (2018) conducted an analysis of traffic flow patterns by fine-tuning pre-trained CNN models using a series of long-time closed-circuit television (CCTV) images. By constructing a spatio-temporal matrix as the input for the CNN model based on traffic data, Zhang *et al.* (2019b) and Ma *et al.* (2017) enabled the accurate predictions of traffic flow and traffic speed, respectively. These studies show that deep learning models can solve complicated problems with high dimensional data over large spatial extents, which become a potent alternative for traffic modeling. However, these estimation models usually require a large volume of traffic noise-related data as well as environmental factors from multiple sources for making estimations, such as pavement material, traffic flow, heavy vehicle ratio, intersection parameters, road network data, digital elevation model (DEM), meteorological features, street trees, and building information. Consequently, the availability of data became an obstacle to the use of these models.

Street view imagery (SVI) is an emerging geospatial big data that depicts urban physical space from human perspectives, including images provided by map service providers and social media photos uploaded by users (Anguelov *et al.* 2010). SVI preserves detailed visual information about street-level facades with a wide geographic coverage and has relatively high acquisition efficiency and cost-effectiveness, which is appropriate for large-scale urban environment auditing (Biljecki and Ito 2021, Fan *et al.* 2023). The use of SVI in auditing urban environments can be classified into two categories: element-level observation and scene-level observation (Kang *et al.* 2020). The former refers to the quantitative analysis of the properties of the objects contained in the street view, including object detection (Frome *et al.* 2009, Flores and Belongie 2010, Yin *et al.* 2015, Zhang *et al.* 2018) and indices calculation (Li *et al.* 2015, Gong *et al.* 2018, Zeng *et al.* 2018, Hu *et al.* 2020). The latter mostly refers to an overall understanding of the entire scene and the underlying semantics, including the service functions, economic conditions, and human subjective perception of urban space (Ye *et al.* 2021, Kang *et al.* 2021, 2023).

Regarding the long-term co-evolutionary process between the built environment and the intensity of human activity, as well as their interactions, it might be possible to estimate human activity intensity from the physical built environment (Zhang *et al.* 2019a). Urban traffic noise serves as a significant indicator of human activity intensity. At the scene level, as a description of real-world scenery, SVI may contain extensive information for understanding urban traffic noise along the road network, including both street configurations and the surrounding physical environment, e.g. paving materials, road width, vehicle types, traffic volume, roadside greening, building heights, density, etc. By using deep learning techniques, it is possible to extract such visual features that reveal the characteristics of acoustic environment for the noise estimation task. Furthermore, street view images systematically cover nearly all levels of the urban road network, and the large geographical coverage and high volume of SVI datasets provide researchers with an unprecedented opportunity for comprehensively modeling urban road traffic noise at a large but fine scale. Zhao *et al*. (2023) conducted an assessment of urban soundscapes utilizing street view imagery and perception-based indicators scored by volunteers. Nonetheless, it is important to acknowledge that their estimation model relies on human perception as a basis. In contrast, a more standardized and objective approach might be achieved through physical measurements based on real-world data.

To this end, we propose a computational framework to estimate traffic noise along the road network using street view images and deep learning approaches based on real-world measurements. Such a framework contains three steps including data collection and processing, DCNN-based features extraction, model training and mapping (via a simple yet efficient data-driven approach, achieve end-to-end output from street view images to road traffic noise (sound perception in decibel (dB) levels), and map the distribution of city-scale road traffic noise to a finer spatial granularity). To the best of our knowledge, this study is among the first to quantitatively estimate road traffic noise in decibel levels by utilizing street view images as the input data source. Using this framework, we can create urban street-level noise maps with low cost, low data needs, and high efficiency, which can be generalized among regions and may be used as reliable and powerful tools to further benefit a series of applications in transportation, social sensing, environmental psychology, and human mobility.

The paper is organized as follows: Section 2 introduces the proposed road traffic noise estimation methodology framework. Subsequently, Section 3 presents an illustrative case study, and in Section 4 we present the main results. Section 5 summarizes the potential application of our work and discusses limitations and future directions. Section 6 gives the concluding remarks. It should be noted that the concepts of 'road traffic noise' and 'noise along road network' are synonymous in the context of this study, as the noise along the road network predominantly arises from traffic and transportation sources.

2. Methodology

Here, we introduce a novel computational framework for estimating traffic noise along the road network, which is substantially validated through a case study in Section 3. As is depicted in Figure 1, our computational framework mainly contains three steps: (1) Data collection and processing. (2) DCNN-based features extraction. (3) Model training and mapping. Specifically, we collect multiple datasets, including road networks, street view images, and in-situ road traffic noises. The next procedure is data processing for training dataset preparation. After that, we train a DCNN-regression model to estimate road traffic noise from street view images (SVIs).

2.1. Noise estimation from SVI

Treating road traffic noise estimation as an end-to-end CNN regression task presents a challenge in enhancing model stability due to its inherent complexity and uncertainty.

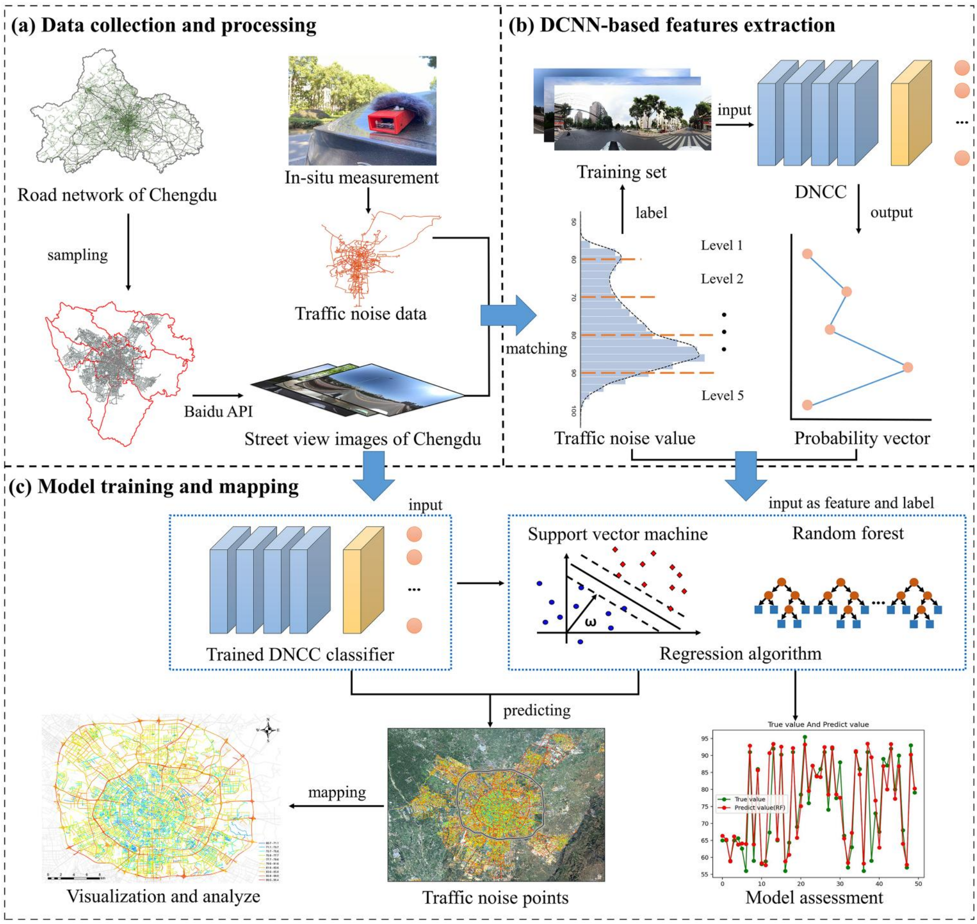

Figure 1. The computational framework for estimating traffic noise along the road network.

Therefore, we adopt a classification-then-regression strategy to estimate traffic noise along the road network from street view imagery in this study. Such a strategy achieves the balance between efficiency and accuracy in prior studies (Wu et al. 2019, Zhang *et al*. 2019a, Ji *et al*. 2021). Moreover, research have shown that roads with similar transportation function tends to exhibit comparable street characteristics (Barrigón Morillas *et al.* 2005, Carmona del Río *et al.* 2011). However, the visual features depicted in street view images demonstrate a wide range of diversities and complexities. To address these issues, our study initially employs a deep convolutional neural network (DCNN) classification technique to measure variances and commonalities within street view images associated with varying noise levels. Then, we convert the output probability vectors back to the noise values through regression algorithms. Utilizing a classification model enhances the capability to capture these features effectively and facilitates the learning of more meaningful representations. Consequently, this approach contributes to the improvement of accuracy and overall performance in the subsequent regression task.

Based on the geographical coordinates, the road traffic noise values from in-situ physical measurements were utilized as the ground truth label to establish the relationship between the two modalities, i.e. street view imagery and sounding information in decibel levels in this work. Subsequently, street view images were converted to N category based on the corresponding continuous noise values $N_{observed}$. Specifically, the SVIs are labeled as the same category if their noise values are similar, ranging from level 1 (low noise) to level n (high noise).

In this study, a DCNN is trained to learn the traffic noise-related characteristics from the road environment. Such a model has demonstrated high performances in prior researches in cross-modal tasks, which proves its strong inference and reasoning capabilities (Wang *et al*. 2016, Wei *et al*. 2017, Nagrani *et al*. 2018). The high dimensional traffic noise-related visual features in SVIs are then extracted using the DCNN, and subsequently undergo a softmax function to obtain the output probability vector. In particular, the output of the model is an n-dimensional vector $Q = (q_1, q_2, q_3, \ldots, q_\mathrm{n})$ that represents the estimation probability of each category that the image may belong to. For instance, if $q_1 = 0.1$, it indicates that there is only a 10% chance that the image belongs to the category 'low noise' (level 1).

Then, we utilize regression algorithms to obtain numerical estimations of road traffic noise in decibels and input the extracted training feature vectors Q and their corresponding labels $N_{observed}$ for training the regression model.

After finishing the training procedure of such a model, all street view images were input into the well-trained model to produce the traffic noise value $N_{predicted}$. Theoretically, such a model can be used for estimating noise values with any given street view image.

2.2. DCNN interpretation

DCNN can extract detailed spatial information in the lower layer and abstract semantic information in the higher layer through consecutive convolutional operations. Although deep learning approaches have achieved satisfactory performance in a wide range of computer vision tasks, they still lack interpretability and are viewed as black boxes.

Class activation mapping (CAM) provides a straightforward visualization approach to meet the increasing demands for better comprehending the internal behavior of a CNN model (Zhou *et al*. 2016). Let f be an original estimation model and c denotes a target class of interest, CAM takes a weighted combination of the activation maps from a convolutional layer to locate the discriminative objects of a specific input x (i.e. $f^c(x)$), then the heat map is produced for showing the relevance of individual pixels for image classification, which can be generated by:

$$L^c_{CAM}(A) = \mathrm{ReLU}\left(\sum_{k=1}^{N_l} \alpha_k A_k\right)$$

with $A = f^{[l]}(x)$, where $f^{[l]}(\cdot)$ denotes the output of the l-th layer. A_k is a k-th activation map of A and α_k is A_k's coefficient (i.e. the importance). N_l indicates the number of the activation maps of the l-th layer.

Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm is a generalized version of CAM, which is applicable to a wide variety of CNN models and can be applied to particular layers of the network. In this paper, we use the Grad-CAM technique to provide a qualitative visual interpretation of estimating road traffic noise levels based on street view images. We use the last convolutional layer of our network as the target layer to generate the activation heat map and superimpose it onto the original input image to achieve better visualization and interpretability. By doing so, the effectiveness of the model can be evaluated.

3. Case study

3.1. Study area

We take Chengdu, located in southwest China and the capital city of Sichuan Province, as our study area. Chengdu has a large permanent population of nearly 21 million residents (2020), and the second-largest vehicle ownership (more than 5,000,000 in 2020) nationwide. We conduct our study mainly inside the municipal district, including the central area of Chengdu that comprises five old districts (i.e. Wuhou, Jinjiang, Chenghua, Qingyang, Jinniu), the suburban area (i.e. Pidu, Wenjiang, Shuangliu, Longquanyi, Qingbaijiang, Xindu, except Xinjin), and two newly developed special districts (i.e. Hi-Tech Industrial Development Zone and Tianfu New Area). The study area together with the spatial distribution of main road types in Chengdu are mapped in Figure 2.

3.2. Data collection

3.2.1. Road network

Open Street Map (OSM) is a publicly accessible and collaborative spatial database of Volunteered Geographic Information (VGI) that provides annotations of land use at the street level (Jokar Arsanjani *et al.* 2015). Several metadata information such as road types and land use are included. In our work, we leveraged the OSM road network data[1] available within the study area. Subsequently, we categorized urban roads into two primary types (i.e. 'Roads' and 'Link roads'). According to the principal tags in OSM, road tags in OSM are used to represent the main roads themselves, while link road tags are used to represent connections between different types of roads. These two categories cover the majority of roads within the urban network. We aggregated the collected road traffic noise data and attached street view images into distinct road segments for analysis.

3.2.2. Street view images

Street view images (SVIs) provide a fine-grained representation of the physical surroundings along nearly every navigable street, which were captured by the cameras mounted on mobile vehicles. Our street view imagery was downloaded from Baidu Map, as Baidu street view (BSV) service has an extensive geographic coverage of the urban areas in China with high quality and fast update frequency. In this work, each panorama image, sized 1024 by 360 pixels, includes a place-ID number and corresponding latitude and longitude coordinates, with azimuth and elevation angles set to 0 degrees.

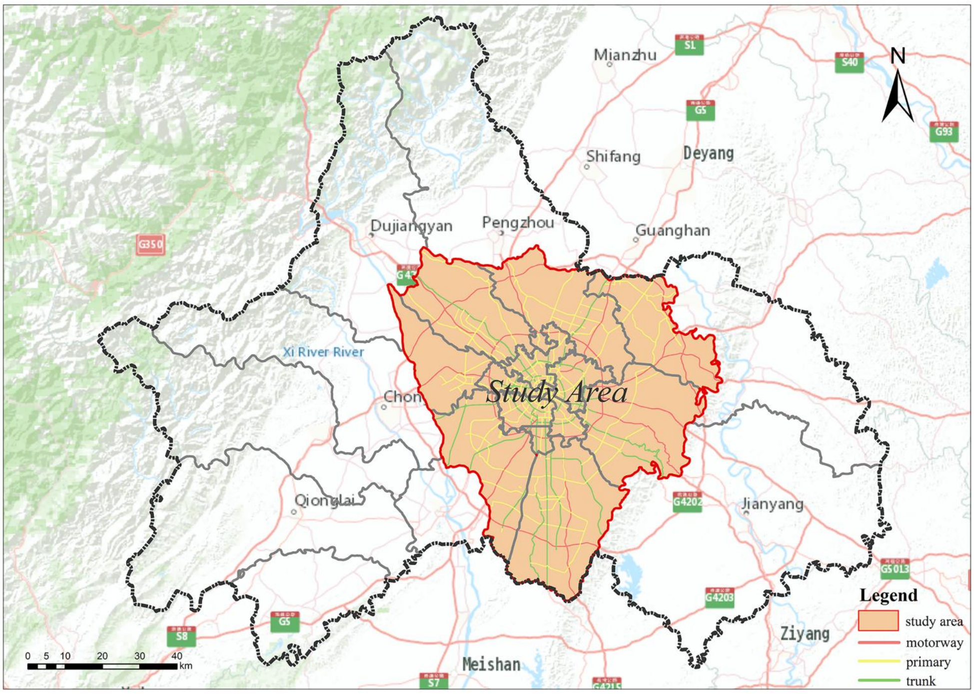

Figure 2. Study area where traffic noise data along the road network were collected.

3.2.3. In-situ road traffic noise along the road network

Road traffic noise data was conventionally collected using noise monitoring stations in prior studies. Recently, researchers have employed floating cars to collect traffic-related data which could produce wide-coverage and high-quality datasets. For example, taxi fleets have been used in several certain applications as floating cars (Brockfeld et al. 2007, Rahmani et al. 2013, Stamos *et al.* 2016). Inspired by them, we designed and mounted a portable device on a moving vehicle to acquire traffic noise along the road network, as opposed to collecting stationary data.

In order to create a portable device for real-time road traffic noise data acquisition, we employed a Raspberry Pi as the foundational hardware platform. The device is contains two key components: the GPS module and the Noise collection module. By doing so, we could synchronize coordinate positioning and noise acquisition, as precise vehicle positioning estimates are essential for associating the traffic noise data along the road network with high-resolution street view panoramas in this work.

The sampling rate of the device is five seconds. For each record, the date, time, longitude, latitude, and in-situ road traffic noise in decibels are included. We later processed the absolute noise value using the moving average method over a time window of 20 seconds. Notably, the interference of noise from other sudden sources, such as the appearance of sirens, was attenuated by averaging the current noise record with the previous 3 records, furthermore, to minimize the effect of random fluctuations during the data collection period, we took multiple samples of most of the road segments at different times. With this device, the field data was collected by moving the experimental vehicle to capture the noise generated primarily by road traffic. Such a data source may yield some other benefits such as high coverage and adequate data volume, cost-efficiency, and timeliness.

In addition, the device was affixed to the boot lid of an electric experimental vehicle to avoid petrol engine vibrations in noise measurement. We acknowledge that wind may influence road traffic noise measurements, especially when the vehicle has a high speed (exceeding 20 km/h). To minimize this effect, both hardware and software measures have been employed. On the hardware side, the entire unit was covered by a microphone windproof fleece; on the software side, we performed detrending of the original data by applying piecewise regression, which will be described in Section 3.3.

3.3. Data processing

3.3.1. Data preprocessing of road traffic noise

To maintain the data quality of the collected traffic noise along the road network and further construct the training dataset, we performed a series of data processing steps, which are outlined below:

1. The driving speed of the experimental vehicle was calculated by determining the time interval between every two nonconsecutive data points in the record sequence and computing their geographic distance via the GPS coordinates.
2. We removed data outliers with the following criteria to reduce potentially erroneous observations: Data points with a speed of 0 km/h or greater than 140 km/h were filtered, and the abnormally high values of road traffic noise data were excluded.
3. To measure the variations in noise level due to changes in the collector car's speed, and to further eliminate noise correlated with the vehicle's speed, we created scatter plots and performed regression analysis between the car's moving speed and the collected noise.
4. In order to eliminate the trend observed in step 3, we opted for a method of piecewise detrending to correct the original data (Seber and Wild 1989), according to their noise decibels. The entire data sequence was divided into 5 subsequences, and each was detrended individually as we will perform a 5-category classification to train a DCNN model.
5. For each geo-referenced traffic noise point along the road network, we matched it to the corresponding OSM road type according to its positional coordinates.

3.3.2. Construct training dataset

Based on the processed road traffic noise data, we retrieved street view images within a 50 m radius of each traffic noise point along the road network. It should be noted that a street view image may be matched to multiple noise data points. To account for this, we computed the average value as the corresponding road traffic noise for each image. Additionally, we excluded panoramas captured in street sites within tunnels. Then, to train the deep learning model for estimating road traffic noise, we adopted the strategy of performing a 5-category classification in which the continuous noise values were converted into five categories. The SVIs of noise values falling in the same range were labeled as the same category, ranging from level 1 to level 5.

Additionally, since the categories display an approximately Gaussian distribution pattern in our study, data augmentation was implemented to address the problem of class imbalance and potential overfitting to the under-represented classes. Specifically, in this work, we mainly applied color enhancement and horizontal flipping to the training samples of the minority classes as they can effectively generate images with realistic illumination changes and better simulate realistic street scenes, and we reconstructed the majority class with a stratified sampling strategy for each road type.

3.4. Noise estimation

3.4.1. Model training

The DCNN algorithm was implemented based on the PyTorch deep learning framework. The model's inputs include street view images and their corresponding noise level labels. The street view imagery dataset was randomly partitioned into a training dataset and a validation dataset, with the former containing 80% of all street view images, which were then fed into the model.

The ResNet network has become one of the most popular and widely used DCNNs at present due to its high performance in a series of tasks in the field of computer vision. In this case, we employed the ResNet (He *et al*. 2016) model with 34 convolutional layers (i.e. ResNet34) for automatic learning road-level acoustic information. To obtain noise values, the Support Vector Machine (SVM) (Cortes and Vapnik 1995) and Random Forest (RF) (Breiman 2001) algorithms, which have been demonstrated to be highly efficient in solving regression problems, were then applied to infer the estimated noise values.

3.4.2. Performance evaluation

To evaluate the performance of our proposed model, the most widely used evaluation metrics were employed in this work. Specifically, we used *Accuracy*, *Precision*, *Recall,* and *F1-score* for the classification task and R-square (R^2), mean absolute error (*MAE*), mean squared error (*MSE*), and root mean square error (*RMSE*) for evaluating the overall performance of the DCNN-Regression model.

4. Results

4.1. Descriptive statistics of collected road traffic noise

A total number of 116,483 road traffic noise data was collected in Chengdu from 17 March to 11 April 2022. The data collection took place from 9 am to 5 pm daily, and the vehicle's effective mileage was recorded as 224,332 meters. The collected data ranged from 51 to 101 dB and approximately exhibited a normal distribution. The dataset's average, median, and standard deviation were 78.5 dB, 81 dB, and 9.95 dB, respectively.

Table 1 illustrates the interval classification of road traffic noise levels and the corresponding number of specific road types associated with the traffic noise data along the road network at each level. It needs to be noted that the unclassified type does not mean an unknown road classification, but rather minor roads of a lower classification than tertiary in the road network system. In our study, road traffic noises ranging from 80 to 85 dB constituted the highest proportion among the majority of road types in the raw data.

4.2. Speed-related noise correction

In Figure 3, a scatter plot is created to display the relationship between the instantaneous speed of the experimental vehicle and the corresponding road traffic noises, along with the regression line. The correlation coefficient obtained from the data was 0.54, indicating a moderate level of positive correlation, which was found to be statistically significant with a *p*-value of less than 0.01. The coefficient of determination suggests that about 30% of the total variation of road traffic noise can be explained by the variation in vehicle speed.

The raw road traffic noise data was partitioned into five subsets based on the noise decibels. As shown in Table 2, we list the correlation coefficient and *y*-intercept of the piecewise linear regression. After the correction procedure for each subset, we classified the road traffic noise into five categories for the DCNN model's input and created kernel density maps for each category. Figure 4 depicts the spatial distribution of the raw data of in-situ noise collection and the processed traffic noise data along the road network in Chengdu city during our collection period, with colored areas representing spatial clusters of dense noise points. As shown in Figure 4, spatial clusters do not significantly appear in level 1.

4.3. Model accuracy assessment

A total of 46,280 images were used in the training set, and 9,256 images were used for validation. Then, a ResNet-34 was trained to infer road traffic noise levels from street view images. The training process employed a cross-entropy loss function. During the training phase, cosine decay is used for learning weight adjustment, and we used SGD as the optimizer with momentum set to 0.9 as well as weight decay (1e-4) to avoid overfitting.After training for 200 epochs, we reset the learning rate to the initial learning rate, and the model weights learned are used as initial weights in the next restart. This training strategy is called SGDR (Loshchilov and Hutter 2016) and performed well in our classification task.

Our classification model achieved an overall accuracy of 75.7% on the validation sets, the *Precision*, *Recall,* and *F1-score* were 0.774, 0.736, and 0.755 respectively. The best-performing class (level 5) achieved an accuracy of 95.5%. On the contrary, class level 3 only achieved an accuracy of 52.5%, and class level 1, level 2, and level 4 achieved

Table 1. The number of the measured noise on different levels of roads and link roads.

		Road traffic noise level				
Road types		Level 1 $\leq$75 dB	Level 2 75–80 dB	Level 3 80–85 dB	Level 4 85–90 dB	Level 5 >90 dB
Roads	Motorway	126	188	286	382	126
	Trunk	1202	1736	4330	4980	1202
	Primary	4370	6340	8290	8662	4370
	Secondary	4520	4164	6324	5966	4520
	Tertiary	9784	5300	8526	6212	9784
	Unclassified	3556	710	1086	738	3556
	Residential	2898	1204	2052	1412	2898
Link roads	Motorway link	44	116	210	142	44
	Trunk link	198	232	532	664	198
	Primary link	84	192	300	230	84
	Secondary link	60	148	104	82	60

Note: In OSM, 'roads' are systematically categorized and annotated based on their grade, function, and distinctive attributes. 'Link roads' serve as connectors or ramps between different major roadways.

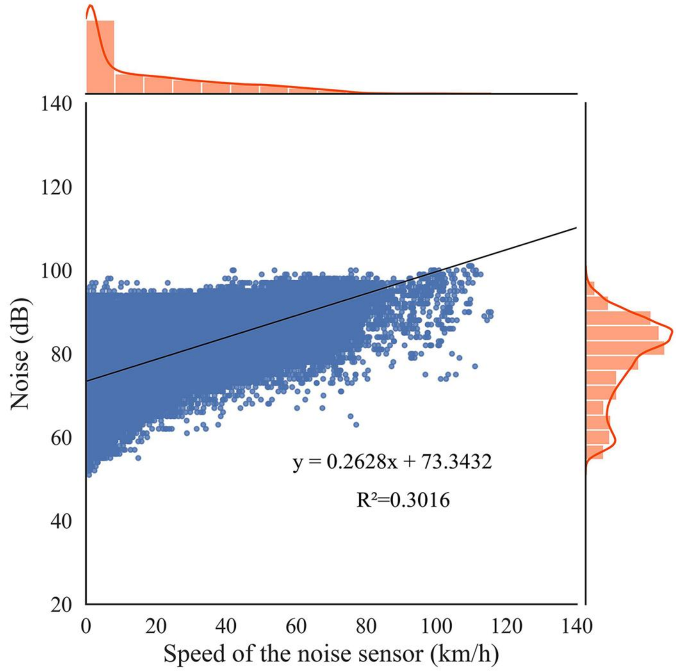

Figure 3. Scatter diagram of the vehicle speed and the measured traffic noise along the road network.

Table 2. Piecewise linear regression.

	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
Noise decibel	≤60	60–70	70–80	80–90	>90
Correlation coefficient	0.06	0.35	0.34	0.17	0.43
y-Intercept	58.21	64.72	74.90	84.65	91.04

accuracies of 79.2%, 83.4%, and 66.9%, respectively. All the results were averaged after cross-validations. The confusion matrix of the DCNN model is shown in Figure 5.

The vector with probability estimations for each class output by the deep learning network and their corresponding noise value were then fed into the regression models. To be specific, SVM with a linear kernel and RF algorithms (hyperparameter set: n_estimators: 300, max_depth: 10, min_samples_leaf: 1, min_samples_leaf: 2, bootstrap: True) were employed for estimating the actual value of road traffic noise. The performance was evaluated using four indicators (R^2, *MAE*, *MSE*, and *RMSE*). The results of the test set of the DCNN-Regression models are provided in Table 3, it is observed that the DCNN-RF exhibits fewer errors and performs slightly better than the DCNN-SVM in our task. The results demonstrate that our procedure for road traffic noise estimation is feasible. In Figure 6, we illustrate several sample street view images of a selected road segment in Chengdu, along with the corresponding road traffic noise values that our model estimated.

4.4. Road traffic noise estimation

By applying such a classification-then-regression strategy, we conducted estimations for a total of 151,747 Baidu street view images collected from both urban and

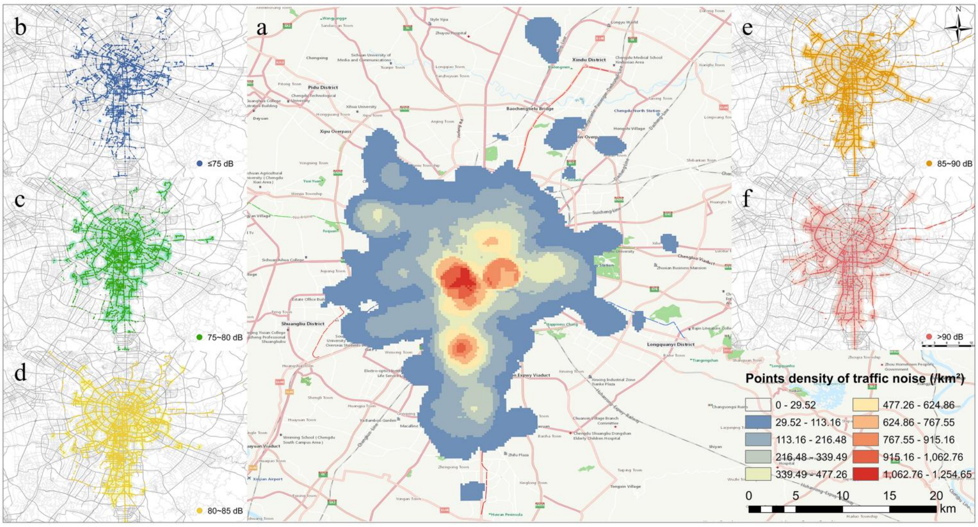

Figure 4. Density map of traffic noise data along the road network. (a) A point density map of in-situ road traffic noise data that we collected within Chengdu. (b–f) Spatial distribution of the processed noise points at each level. (b) level 1 ($\leq$75 dB), (c) level 2 (75–80 dB), (d) level 3 (80–85 dB), (e) level 4 (85–90 dB), (f) level 5 ($>$90 dB).

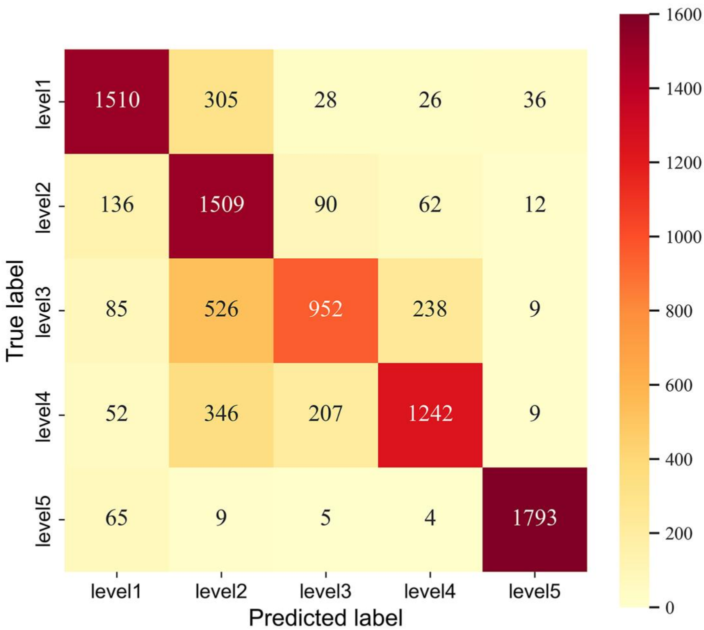

Figure 5. Confusion matrix of the trained network.

suburban areas of Chengdu City (Figure A1(a)). Our analysis focused on the core area, which corresponds to the region within the ring freeway of Chengdu. The study area was initially divided into small square regions of 200 × 200 m using a rectangular grid. The grid map was then obtained by calculating the average values of all the noise points for each grid (Figure A1(b)).

Table 3. Performance of the DCNN-Regression models on the test set with the best performing model represented in bold.

	R^2	*MAE*	*MSE*	*RMSE*
DCNN-SVM	0.61	2.04	8.44	2.91
DCNN-RF	**0.64**	**2.01**	**7.32**	**2.71**

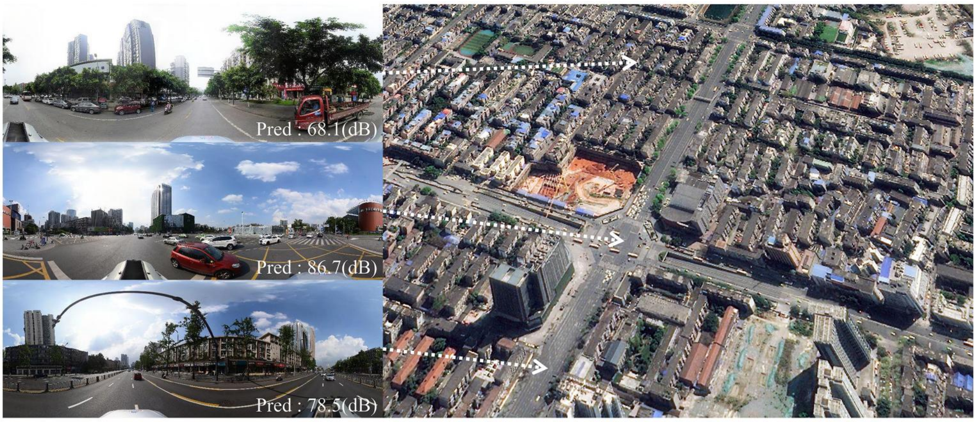

Figure 6. Examples of the SVI-based road traffic noise estimation of a road section in Chengdu. The street view images were sourced from Baidu Map Open Platform (https://lbsyun.baidu.com) and the base map depicting the landscape of a road section in Chengdu was obtained from the website https://siweiearth.com.

To further analyze the estimation results, the OSM roads served as the unit to calculate the average road traffic noise by aggregating all the point-level SVI data into their corresponding road segments. The estimated traffic noise of roads is displayed in Figure 7. Additionally, we generated a plot illustrating the difference between predicted and observed traffic noise values (in decibels) at the road level to visualize which road segments were overrated or underrated. The results are presented specifically within the areas of the third ring road in Figure A2.

According to Figure 7, the two-ring road viaduct, the third-ring road, and the outer ring freeway are found to have higher road traffic noise. Also, noise near the interchanges and signalized intersections reaches relatively high levels. In the downtown area of the city, the average noise levels are relatively low, especially for roads that connect human settlements.

A box plot was created to provide a summary of the traffic noise levels for different road categories, with a particular focus on roads and link roads, which are considered to be dominating contributors to environmental noises. As shown in Figure 8, the noise value of the link road category is distributed within a higher range than that of the most road category, which fits our expectations. Furthermore, lower-grade roads tend to exhibit higher standard deviations in terms of their traffic noises, and there is a noticeable decrease in noise levels from highways to minor roads (the mean value of noise of secondary links is higher than that of primary links).

Figure 7. Estimation of road traffic noise within the outer ring freeway of Chengdu.

4.5. Interpretation of the DCNN model

For better visualization and interpretation, we overlaid the activation heat map onto the original input image. Figure 9 presents the Grad-CAM results of five correctly classified street view images corresponding to each category of road traffic noise from level 1 to level 5, the respective color of the Grad-CAM represents the importance of the network's estimation. For each category, we selected a variety of typical street view images, depicting diverse street scenes from different road types to better understand the DCNN network's attention.

Regarding the street view images that correspond to the 'relatively quiet place' categories (level 1 and level 2), it appears that the DCNN model relies heavily on visual cues related to the surrounding urban environment along the road, including elements such as greenery and residential housing. While for those street view images that were classified as level 3 and level 4, facade structures (infrastructure such as bridges, and buildings), as well as contextual information such as roadside green belts, tend to be emphasized, given that most images were classified into level 3 and level 4, and the corresponding matched street scenes were relatively diverse and complex, no prominent features determine their classification. As shown in class level 5, the classification model mainly focuses on the road areas and the horizontal surroundings of roadways. Particularly, for certain street view images that were classified as high-level road traffic noise, we observed that the sky was highlighted. It indicated that higher noise may occur on roads with a more open line of sight, such as highways, where sky visibility is higher.

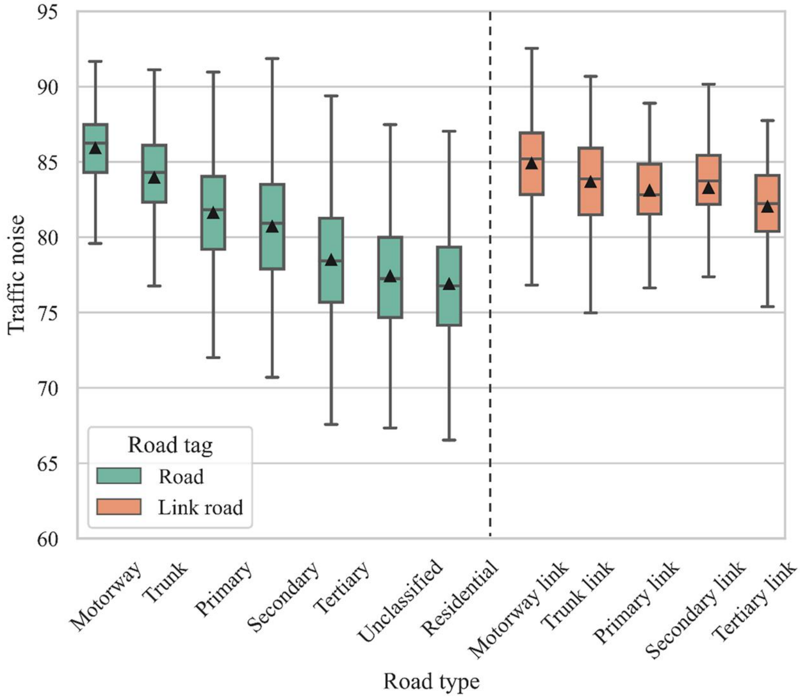

Figure 8. Box plot of noise along the principal local road and the link road with a dotted line divides the two types.

In addition, the analysis of the Grad-CAM results of misclassified samples plays an important role in understanding the reasons behind incorrect predictions made by the model. It aids in whether the model attends to important features or emphasizes irrelevant regions, shedding light on potential factors contributing to the incorrect predictions. Therefore, we selected a representative set of misclassified samples and employed the Grad-CAM technique to generate corresponding visualizations. These visualizations, along with detailed information, are presented in Figure A3. In this task, misclassification can potentially be attributed to similarities between categories and local information misdirection.

5. Discussion

5.1. Discussion of results

According to the frequency distribution of the estimated road traffic values depicted in Figure 10, both the extremely low values and high values are relatively few, and most are under moderate noise levels, this aligns with the expected normal distribution and is consistent with previous road traffic noise monitoring studies, which demonstrate the noise sample data we collected is typical and convincible.

In Section 4.3, the confusion matrix reveals the challenges faced by the deep learning model in distinguishing images classified as level 3 or level 4 noise levels (Figure 5). This suggests a relatively high degree of similarity between the two classes.

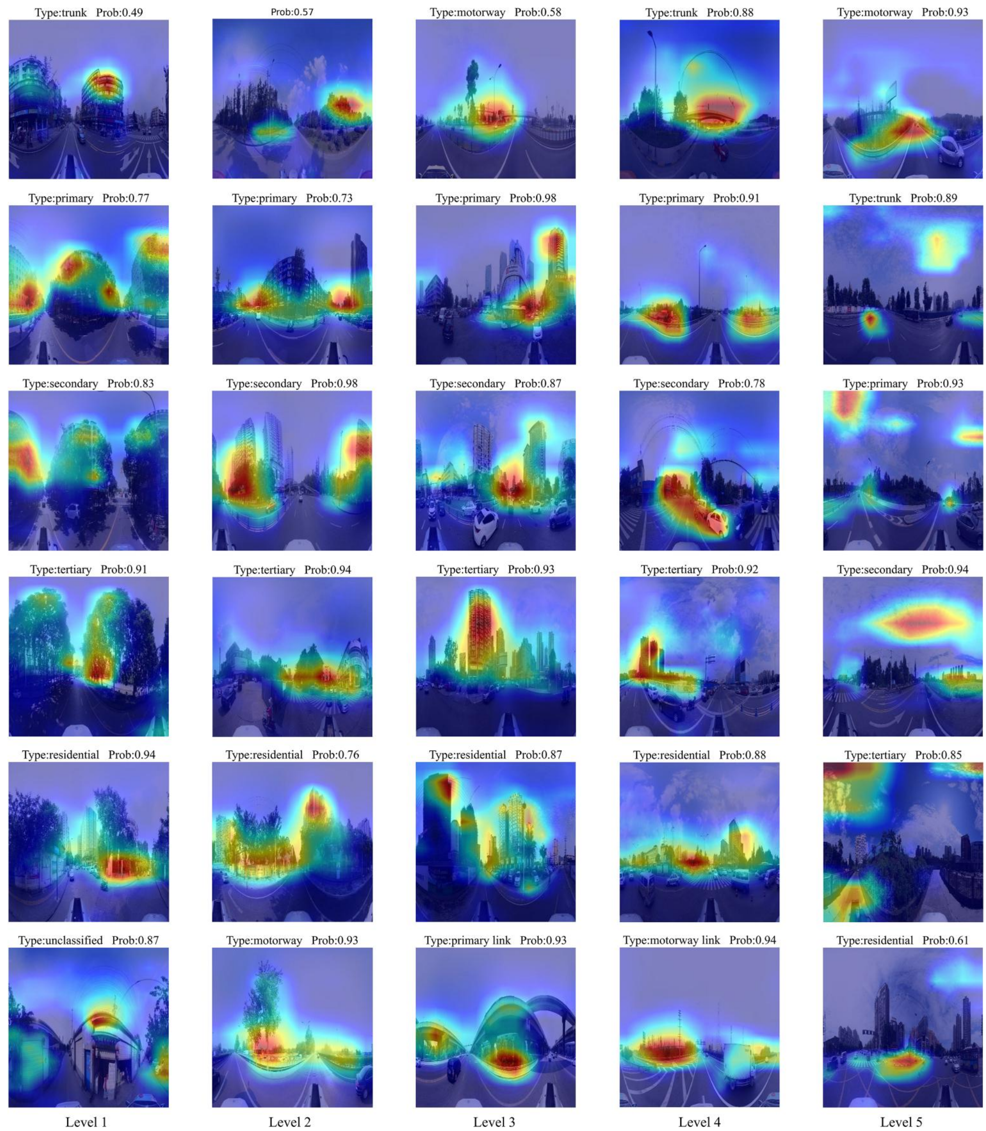

Figure 9. Examples of the correctly classified street view images and their estimated probability with the corresponding Grad-CAMs overlaid.

One possible explanation is that street view images with moderate noise levels tend to have complex traffic scenarios and road layouts. In contrast, street view images with level 1 and level 5 noise levels exhibit relatively low intra-class variance, which may make them easier to differentiate.

By examining relevant street view images in which road traffic noises were either overestimated or underestimated with significant estimation errors (Figure A2) and combining the generated Grad-CAM maps (Figure A3), we observe that the underestimation of noise level may potentially be attributed to the inadequate representation of the acoustic environment by the visual context within the street scene. One potential influencing factor to this discrepancy might be the obscuring or masking of traffic noise sources by visually prominent

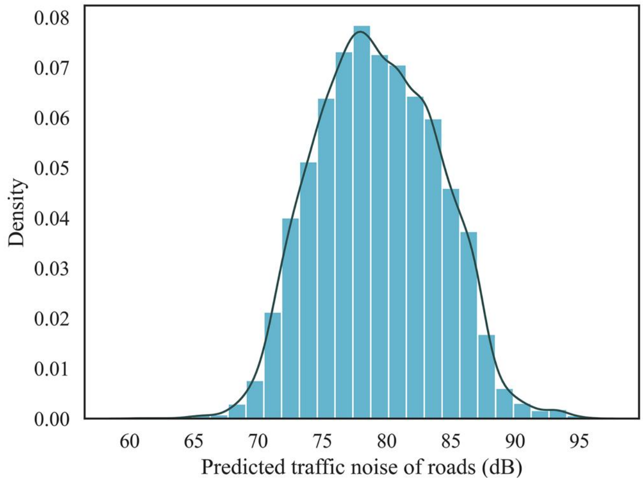

Figure 10. Frequency distribution of estimated road traffic noise within the urban area of Chengdu.

entities such as building facades or trees in the street view imagery. This is also related to the perspective or viewing angle from which the street view image was originally captured.

Nonetheless, focusing on the static attributes of visual elements while ignoring temporal fluctuations in traffic noise may also lead to inaccurate estimations of traffic noise levels, which typically results in overestimations in this study. In road segments with consistently high traffic volumes, such as intersections, significant reductions in traffic noise level may occur during specific time periods. If some of our noise collection time window happens to fall within this period, our model is likely to predict instantaneous noise levels that are higher than the 'ground truth' at these points. It suggests that future studies should incorporate information about the variance in noise levels.

According to the spatial patterns of traffic noise along the road networks, we found that noise pollution is prevalent and serious in the circumferential ring roads and a series of connected roads. A significant number of roads with an average traffic decibel value exceeding 85 were concentrated on the third ring road and the outer ring freeway, which were constructed to accommodate the heavy traffic in the urban transportation system. The high noise emissions on these ring roads and freeways can be ascribed to the large traffic volumes and higher vehicle speeds (with speed limits over 80 km/h). Conversely, subsidiary streets within built-up areas exhibit predominantly lower levels of road traffic noise, as they generally have lower speed limits (ranging from 30-60 km/h) and lower traffic densities.

In addition, we investigate the impacts of land use types on road traffic noise pollution for certain noise-sensitive areas. We obtained the land use data from OSM and overlaid the estimated road noise map on it to examine the noise pollution near different land-use types (Figure 11). We selected some typical sites and presented them in Figure 12.

As shown in Figure 12(a), roads around schools and colleges have relatively high estimated noise levels due to the high volumes of traffic flows. Whereas regions inside these places are quieter, likely due to the legal priority of pedestrians and the prohibition of honking and speeding in such areas. From Figure 12(b), we noticed that residential areas along the main road or the intersections tend to have a considerably higher noise exposure. A potential solution is to install noise barriers at these places to reduce the noise level caused by passing-by traffic flow. Besides, we found that urban gardens and parks are generally associated with lower noise emissions, possibly due to the sound-absorbing effect of plantings (Figure 12(c)), as suggested by existing literature (Fang and Ling 2003, Onder and Kocbeker 2012, Van Renterghem *et al*. 2012, Ow and Ghosh 2017).

5.2. Contributions and potential applications

Our work further extends the applications of street view imagery in urban traffic studies, as well as the concept of 'urban visual intelligence' (Fan *et al*. 2023), and enables the study of the connection between the visible physical environment and acoustic information. In contrast to the existing statistical and computational models, our proposed framework provides a new perspective for estimating road traffic noise, which is effective towards various local conditions and diverse traffic scenarios without requiring specific factors related to sound generation and propagation. Therefore, SVIs-based road traffic noise estimation makes it possible to measure acoustic information at large scales, even in underdeveloped areas. In this paper, we developed a full-coverage distribution map of traffic noise along the road networks in Chengdu city as a case study. The maps might be utilized in a variety of ways to benefit over 20 million residents.

For academic research, the map can be used to identify hot spots of road traffic noise and investigate its relationship with other traffic-related factors such as air pollution, congestion, and road accidents. Moreover, through the integration of city-scale road traffic noise information with other pertinent environmental indicators, such as the air pollution rate and green space coverage, researchers can conduct a comprehensive assessment of urban

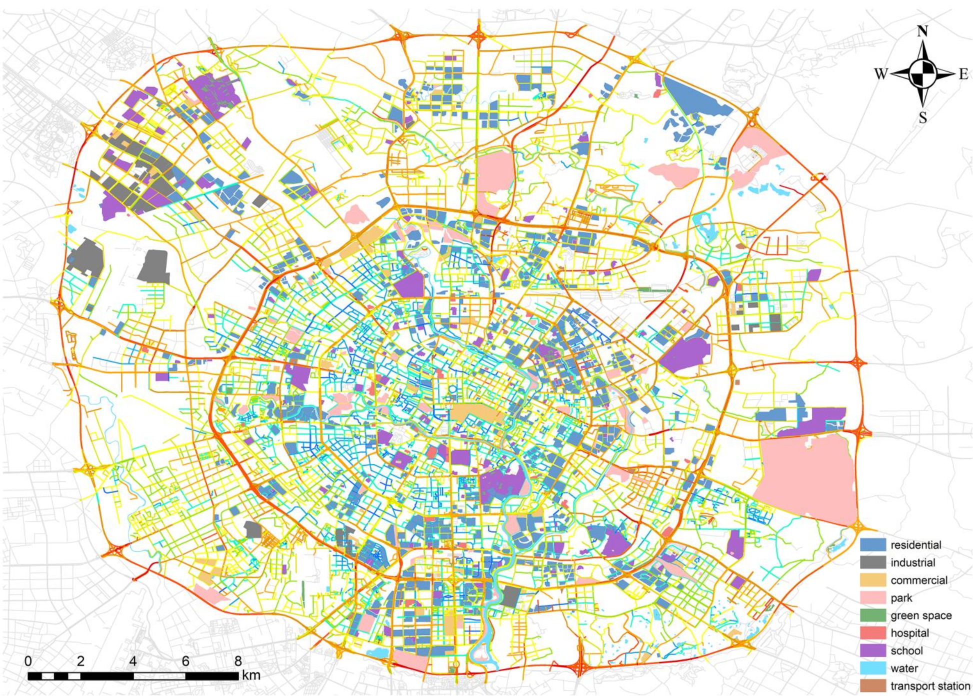

Figure 11. Analysis of road traffic noise sensitive areas within the outer ring freeway of Chengdu.

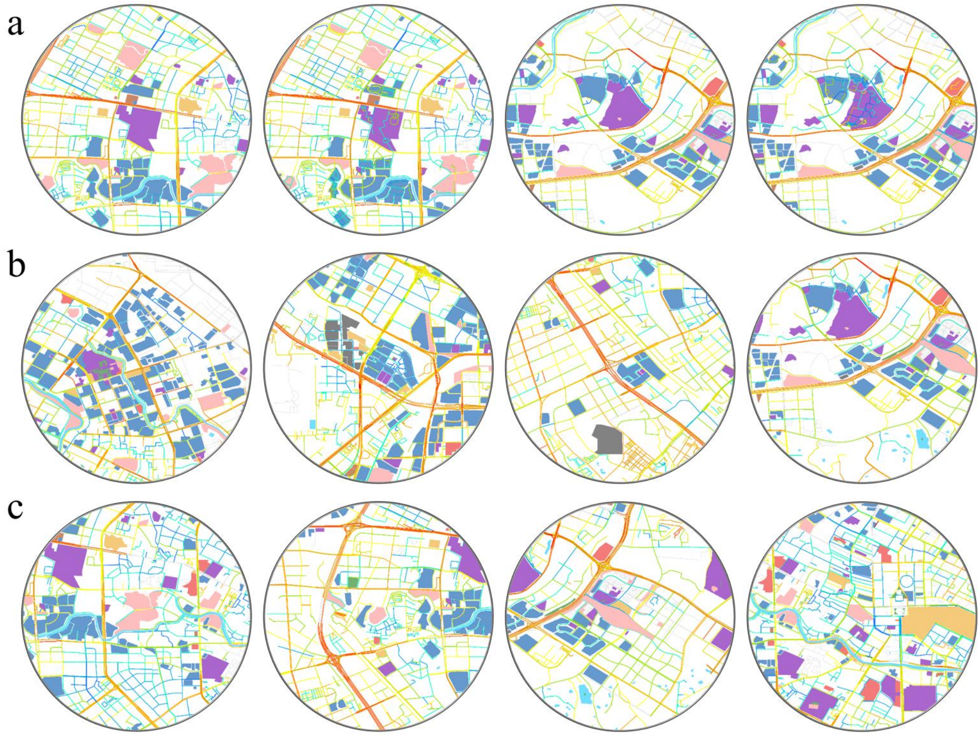

Figure 12. Road traffic noise near the places of schools (a), residential areas (b), and city parks (c).

environmental quality. Furthermore, our research has the potential to complement studies focusing on modeling noise levels on building surfaces along road networks (Zhong *et al.* 2021), which holds promise in facilitating further city investigations and constructions.

For urban planners, the noise map can help estimate the quality of the acoustic environment and inform city policy implementation to minimize and mitigate potential noise sources in sensitive areas like hospitals, schools, and residential communities. As researchers have made efforts in investigating the feasibility of the current constructions of street noise barriers in cities (Qian *et al.* 2022, Zhang *et al.* 2022), the spatial noise map aids in prioritizing areas requiring noise barriers. In addition, it allows for the evaluation of the performance of existing sound barriers. Through a comparative analysis of the noise level portrayed on the map with actual measurement data, areas susceptible to noise leakage or inadequate protection can be discerned, enabling prompt modifications to the design and positioning of sound barriers to improve their efficacy.

5.3. Limitation and future work

This study presents a preliminary investigation of the use of SVIs-based road traffic noise estimation, which has certain limitations that need attention in future work.

It should be acknowledged that this research relies on time-efficient street view images and road network data. Consequently, incorporating more up-to-date data sources is expected to positively impact the predictability and reliability of the research outcomes. Additionally, the collected road traffic noises from our portable device may also contain certain noises such as engine noise, tire rolling noise, and wind noise. Though we have made efforts to mitigate such external sound impacts by physical measures (employed an electric car as our experimental vehicle and used windproof fleece to cover the microphone) as well as a detrending algorithm, it remains challenging to fully eliminate such noises which may influence the accuracy of the results. Thus, there is room to further improve the accuracy and efficiency of noise collection. Moreover, during the data acquisition periods, although we kept the speed of the vehicle as consistent as possible, it may spend more time in a specific area due to reduced speed or traffic congestion, resulting in more frequent data points being captured within the sliding window. Conversely, in free-flowing traffic or at high speeds, vehicle covered larger distances within the same time frame. This can result in sparser sampling as fewer data points were captured. The lower sampling density in these conditions may present challenges in accurately predicting the traffic noise of corresponding road sections.

As evidenced by prior studies, road traffic noise presents distinct patterns during different periods, while we have only considered day time in this research stage. In the future, we are willing to refine our work by exploring the temporal pattern of road traffic noise. For example, noises at night are generally considered more likely to impair residents' physical and mental health, such as annoyance and sleep disorders with awakening. In addition to investigating the periodicity of road traffic noise, such as morning and evening periods, peak and off-peak hours, and weekdays and holidays, it is crucial to recognize that road traffic noise also exhibits continuous variations over time. Consequently, a comprehensive examination of the temporal heterogeneity of road traffic noise becomes imperative in our research efforts.

In addition, the characteristics of road traffic noise may vary among cities or regions due to spatial heterogeneity. Therefore, it is necessary to conduct comparative experiments in multiple regions considering diverse traffic conditions in the future.

6. Conclusion

In this study, we developed an innovative framework to address challenges in collecting traffic noise-related dataset and presented a novel approach for estimating road traffic noise at the city scale. By leveraging large-scale street view imagery and in-situ road traffic noise collected using a custom portable device, we constructed and trained a model capable of estimating acoustic information from street scenes automatically.

In our case study conducted in Chengdu City, we employed a classification-then-regression approach using ResNet and RF regression to balance computational efficiency and predictive accuracy. The corresponding *MAE* and *RMSE* of the DCNN-RF model are 2.01 and 2.71, respectively. Additionally, we utilized the Grad-CAM technique to interpret the DCNN classifier, highlighting the visual cues within the built environment that contribute to the models decision-making. The results revealed that elements related to greenness, openness, building coverage, and road area ratio have the potential to enhance our understanding of traffic-related noise in the urban environment.

The road traffic noise map of our study provides valuable insights for both researchers and policymakers, offering an overview of daytime road traffic noise distribution in urban areas. Additionally, our work investigates deriving sound-related insights from visual features in images, contributing to cross-modal studies in geography and urban planning.

Acknowledgment

The authors would like to thank Urli for the valuable advice provided during the initial stages of the experiment and Mr. Mengze Gao for designing and 3D-printing the enclosure for the data acquisition device. Thanks to the financial support from the National Natural Science Foundation of China [42271476]; the Wuhan University 351 Talent Program, 2020; the State Key Laboratory of Resources and Environmental Information System [2023OPEN007] and the Guangdong Science and Technology Strategic Innovation Fund [2020B1212030009].

Disclosure statement

No potential conflict of interest was reported by the author(s).

Note

1. https://www.openstreetmap.org/

Funding

This work was supported by the National Natural Science Foundation of China [42271476]; the Wuhan University 351 Talent Program, 2020; and the Guangdong Science and Technology Strategic Innovation Fund [2020B1212030009]. This work was also supported by State Key Laboratory of Resources and Environmental Information System [2023OPEN007].

Notes on contributors

Jing Huang is a Master's student at the School of Resource and Environment Sciences, Wuhan University, China. Her research focuses on the analysis of spatio-temporal data in urban geography. Her contributions to the paper include developing traffic noise estimation model, algorithm implementation, conducting case studies, and manuscript writing of this paper.

Teng Fei is an Associate Professor of Cartography and GIScience at the School of Resources and Environment Science, Wuhan University, specializing in the study of urban geographic big data and ecological remote sensing. He contributed to the ideation, conceptualizing, the design of a portable device for in-situ traffic noise data acquisition and manuscript revision.

Yuhao Kang is an assistant professor in GIScience directing the GISense Lab at the Department of Geography, University of South Carolina. His research interests include Human-centered Geospatial Data Science, GIScience, GeoAI, and Urban Visual Intelligence. He contributed to the development of the methodology, as well as the review and editing of this manuscript.

Jun Li graduated from the School of Resource and Environment Sciences, Wuhan University, China. His research is oriented toward geospatial analysis. He contributed to the data processing of street view imagery and traffic noise data.

Ziyu Liu graduated from the School of Resource and Environment Sciences, Wuhan University, China. Her recent work focuses on the use of accurate road PV production estimation from street view image. She contributed to the data collection and curation of the street view images in this work.

Guofeng Wu is a Professor at the Department of Urban Informatics, Shenzhen University, China. His research focuses on the application of remote sensing to natural resources and ecological environments. He is co-responsible for the presentation of this paper.

ORCID

Teng Fei http://orcid.org/0000-0002-3415-1654
Yuhao Kang http://orcid.org/0000-0003-3810-9450

Data and codes availability statement

The original in-situ road traffic noise data with geographic coordinates collected by the experimental vehicle using our portable device, as well as the road traffic noise estimation model and some sample street view images used for demonstration are available at https://github.com/kellyhuang313/traffic-noise-estimation. Instructions for executing the code are provided in the README.txt.

References

Al-Ghonamy, A., 2009. Analysis and evaluation of road traffic noise in Al-Dammam: a business city of the eastern province of KSA. *Journal of Environmental Science and Technology*, 3 (1), 47–55.

Alam, P., *et al.*, 2020. Noise monitoring, mapping, and modelling studies–a review. *Journal of Ecological Engineering*, 21 (4), 82–93.

Anguelov, D., *et al.*, 2010. Google street view: capturing the world at street level. *Computer Magazine.*, 43 (6), 32–38.

Asfiati, S., *et al.*, 2020. Measurement and evaluation of sound intensity at the Medan Railway Station using a sound level meter. ed. *Journal of Physics: Conference Series*, 1428 (1), 012063.

Aumond, P., *et al.*, 2018. Kriging-based spatial interpolation from measurements for sound level mapping in urban areas. *The Journal of the Acoustical Society of America*, 143 (5), 2847–2857.

Banerjee, D., *et al.*, 2008. Evaluation and analysis of road traffic noise in Asansol: an industrial town of eastern India. *International Journal of Environmental Research and Public Health*, 5 (3), 165–171.

Barrigón Morillas, J.M., *et al.*, 2005. A categorization method applied to the study of urban road traffic noise. *The Journal of the Acoustical Society of America*, 117 (5), 2844–2852.

Barry, T.M., and Reagan, J.A., 1978. *FHWA highway traffic noise prediction model*. Washington, DC: Department of Transportation.

Biljecki, F., and Ito, K., 2021. Street view imagery in urban analytics and GIS: a review. *Landscape and Urban Planning*, 215, 104217.

Brambilla, G., *et al.*, 2020. Classification of urban road traffic noise based on sound energy and eventfulness indicators. *Applied Sciences*, 10 (7), 2451.

Bravo-Moncayo, L., *et al.*, 2019. A machine learning approach for traffic-noise annoyance assessment. *Applied Acoustics*, 156, 262–270.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32.

Brockfeld, E., *et al.*, 2007. Benefits and limits of recent floating car data technology–an evaluation study. *In: Proceedings of the 11th WCTR conference*, Berkeley, CA, 24–28.

Can, A., Dekoninck, L., and Botteldooren, D., 2014. Measurement network for urban noise assessment: Comparison of mobile measurements and spatial interpolation approaches. *Applied Acoustics*, 83, 32–39.

Carmona del Río, F.J., *et al.*, 2011. A street categorization method to study urban noise: The Valladolid (Spain) study. *Environmental Engineering Science*, 28 (11), 811–817.

Chen, L., *et al.*, 2021. Modelling traffic noise in a wide gradient interval using artificial neural networks. *Environmental Technology*, 42 (23), 3561–3571.

Cohen, B.S., *et al.*, 2008. Airport-related air pollution and noise. *Journal of Occupational and Environmental Hygiene*, 5 (2), 119–129.

Cortes, C., and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20 (3), 273–297.

Doygun, H., and Kuşat Gurun, D., 2008. Analysing and mapping spatial and temporal dynamics of urban traffic noise pollution: a case study in Kahramanmaraş, Turkey. *Environmental Monitoring and Assessment*, 142 (1–3), 65–72.

Fan, Z., *et al.*, 2023. Urban visual intelligence: uncovering hidden city profiles with street view images. *Proceedings of the National Academy of Sciences of the United States of America*, 120 (27), e2220417120.

Fang, C.-F., and Ling, D.-L., 2003. Investigation of the noise reduction provided by tree belts. *Landscape and Urban Planning*, 63 (4), 187–195.

Flores, A. and Belongie, S., 2010. Removing pedestrians from google street view images. In: *2010 IEEE computer society conference on computer vision and pattern recognition-workshops*. New York, NY: IEEE, 53–58.

Freitas, E., *et al.*, 2012. Traffic noise abatement: How different pavements, vehicle speeds and traffic densities affect annoyance levels. *Transportation Research Part D: Transport and Environment*, 17 (4), 321–326.

Frome, A., *et al.*, 2009. Large-scale privacy protection in google street view. In: *2009 IEEE 12th International Conference on Computer Vision*. Kyoto, Japan: IEEE, 2373–2380.

Gao, W., *et al.*, 2015. Assessment on road traffic noise in the Xizhimen overpass area. *In: Proceedings of the 15th COTA international conference of transportation professionals*. Beijing, China: American Society of Civil Engineers (ASCE), 3290–3299.

Garg, N., and Maji, S., 2014. A critical review of principal traffic noise models: strategies and implications. *Environmental Impact Assessment Review*, 46, 68–81.

Gershon, R.R., *et al.*, 2006. Pilot survey of subway and bus stop noise levels. *Journal of Urban Health* , 83 (5), 802–812.

Ghotbi, M.R., *et al.*, 2012. Noise pollution survey of a two-storey intersection station in Tehran metropolitan subway system. *Environmental Monitoring and Assessment*, 184 (2), 1097–1106.

Givargis, S., and Mahmoodi, M., 2008. Converting the UK calculation of road traffic noise (CORTN) to a model capable of calculating LAeq, 1h for the Tehran's roads. *Applied Acoustics*, 69 (11), 1108–1113.

Gong, F.-Y., *et al.*, 2018. Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Building and Environment*, 134, 155–167.

Hamad, K., Khalil, M.A., and Shanableh, A., 2017. Modeling roadway traffic noise in a hot climate using artificial neural networks. *Transportation Research Part D: Transport and Environment*, 53, 161–177.

Harman, B.I., Koseoglu, H., and Yigit, C.O., 2016. Performance evaluation of IDW, Kriging and multiquadric interpolation methods in producing noise mapping: A case study at the city of Isparta, Turkey. *Applied Acoustics*, 112, 147–157.

He, K., *et al.*, 2016. Deep residual learning for image recognition. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, NV: IEEE, 770–778.

Heutschi, K., 2004. SonRoad: New Swiss road traffic noise model. *Acta Acustica United with Acustica*, 90 (3), 548–554.

Hu, C.-B., *et al.*, 2020. Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. *Building and Environment*, 167, 106424.

Ji, H., *et al.*, 2021. A new data-enabled intelligence framework for evaluating urban space perception. *ISPRS International Journal of Geo-Information*, 10 (6), 400.

Jokar Arsanjani, J., *et al.*, 2015. An introduction to OpenStreetMap in geographic information science: experiences, research, and applications. *In:* J. Jokar Arsanjani, *et al.* eds. *OpenStreetMap in GIScience*. Cham, Switzerland: Springer, 1–15.

Kang, Y., *et al.*, 2020. A review of urban physical environment sensing using street view imagery in public health studies. *Annals of GIS*, 26 (3), 261–275.

Kang, Y., *et al.*, 2021. Human settlement value assessment from a place perspective: Considering human dynamics and perceptions in house price modeling. *Cities*, 118, 103333.

Kang, Y., *et al.*, 2023. Assessing differences in safety perceptions using GeoAI and survey across neighbourhoods in Stockholm, Sweden. *Landscape and Urban Planning*, 236, 104768.
Khan, J., *et al.*, 2018. Road traffic air and noise pollution exposure assessment–a review of tools and techniques. *The Science of the Total Environment*, 634, 661–676.
Kim, M., *et al.*, 2012. Road traffic noise: annoyance, sleep disturbance, and public health implications. *American Journal of Preventive Medicine*, 43 (4), 353–360.
Li, B., Tao, S., and Dawson, R., 2002. Evaluation and analysis of traffic noise from the main urban roads in Beijing. *Applied Acoustics*, 63 (10), 1137–1142.
Li, X., *et al.*, 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14 (3), 675–685.
Loshchilov, I., and Hutter, F., 2016. Sgdr: stochastic gradient descent with warm restarts. arXiv Preprint arXiv:1608.03983. https://doi.org/10.48550/arXiv.1608.03983.
Ma, X., *et al.*, 2017. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17 (4), 818.
Mansouri, N., Pourmahabadian, M., and Ghasemkhani, M., 2006. Road traffic noise in downtown area of Tehran. *Journal of Environmental Health Science & Engineering*, 3 (4), 267–272.
Mehdi, M.R., *et al.*, 2011. Spatio-temporal patterns of road traffic noise pollution in Karachi, Pakistan. *Environment International*, 37 (1), 97–104.
Murphy, E., and King, E.A., 2010. Strategic environmental noise mapping: methodological issues concerning the implementation of the EU environmental noise directive and their policy implications. *Environment International*, 36 (3), 290–298.
Nagrani, A., Albanie, S., and Zisserman, A., 2018. Seeing voices and hearing faces: Cross-modal biometric matching. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. Salt Lake City, UT: IEEE, 8427–8436.
Nedic, V., *et al.*, 2014. Comparison of classical statistical methods and artificial neural network in traffic noise prediction. *Environmental Impact Assessment Review*, 49, 24–30.
Neitzel, R., *et al.*, 2009. Noise levels associated with New York City's mass transit systems. *American Journal of Public Health*, 99 (8), 1393–1399.
Okonofua, E., Ehigiator-Irughe, R., and Ekun, M., 2016. Traffic noise monitoring at road intersections in urban settlement: case study of Ramat Park Benin City. *Nigerian Journal of Technology*, 35 (4), 713–717.
Onder, S., and Kocbeker, Z., 2012. Importance of the green belts to reduce noise pollution and determination of roadside noise reduction effectiveness of bushes in Konya, Turkey. *International Journal of Agricultural and Biosystems Engineering*, 6 (6), 373–376.
Ow, L.F., and Ghosh, S., 2017. Urban cities and road traffic noise: reduction through vegetation. *Applied Acoustics*, 120, 15–20.
Paiva, K.M., Cardoso, M.R.A., and Zannin, P.H.T., 2019. Exposure to road traffic noise: annoyance, perception and associated factors among Brazil's adult population. *The Science of the Total Environment*, 650 (Pt 1), 978–986.
Peppa, M., *et al.*, 2018. Urban traffic flow analysis based on deep learning car detection from CCTV image series. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 499–506.
Qian, Z., *et al.*, 2022. Vectorized dataset of roadside noise barriers in China using street view imagery. *Earth System Science Data*, 14 (9), 4057–4076.
Quartieri, J., *et al.*, 2009, December. A review of traffic noise predictive models. *Recent advances in applied and theoretical mechanics, 5th WSEAS International Conference on Applied and Theoretical Mechanics (MECHANICS'09),* Puerto De La Cruz, Tenerife, Canary Islands, Spain, 14–16.
Rahmani, M., Jenelius, E., and Koutsopoulos, H.N., 2013. Route travel time estimation using low-frequency floating car data. *In: Proceedings of the 16th international IEEE conference on intelligent transportation systems (ITSC 2013)*. Piscataway, NJ: IEEE, 2292–2297.
Seber, G., and Wild, C., 1989. *Nonlinear regression*. New York: Wiley, 768.
Singh, D., *et al.*, 2021. Application of machine learning to include honking effect in vehicular traffic noise prediction. *Applied Sciences*, 11 (13), 6030.

Singh, N., and Davar, S.C., 2004. Noise pollution-sources, effects and control. *Journal of Human Ecology*, 16 (3), 181–187.

Son, H.-J., *et al.*, 2014. A study of traffic noise characteristic of pavement types using NCPX method. *In: INTER-NOISE and NOISE-CON congress and conference proceedings*. Melbourne, Australia: Institute of Noise Control Engineering, 3313–3318.

Stamos, I., *et al.*, 2016. Modeling effects of precipitation on vehicle speed: floating car data approach. *Transportation Research Record: Journal of the Transportation Research Board*, 2551 (1), 100–110.

Tah-Chew, T., and Keung, W.C., 1991. Noise surveys of hand-held pneumatic rock drills in Hong Kong. *Environmental Monitoring and Assessment*, 19 (1–3), 519–523.

Talotte, C., *et al.*, 2003. Identification, modelling and reduction potential of railway noise sources: a critical survey. *Journal of Sound and Vibration*, 267 (3), 447–468.

Trombetta Zannin, P.H., and Bunn, F., 2014. Noise annoyance through railway traffic-a case study. *Journal of Environmental Health Science & Engineering*, 12 (1), 14.

Van Renterghem, T., Botteldooren, D., and Verheyen, K., 2012. Road traffic noise shielding by vegetation belts of limited depth. *Journal of Sound and Vibration*, 331 (10), 2404–2425.

Wallis, A.D., 1997. Airport noise monitoring-a review. *Aircraft Engineering and Aerospace Technology*, 69 (2), 112–118.

Wang, K., *et al.*, 2016. A comprehensive survey on cross-modal retrieval. arXiv Preprint arXiv: 1607.06215. https://doi.org/10.48550/arXiv.1607.06215.

Watts, G., 2005. *Harmonoise prediction model for road traffic noise*. Wokingham: TRL.

Wei, Y., *et al.*, 2017. Cross-modal retrieval with CNN visual features: a new baseline. *IEEE Transactions on Cybernetics*, 47 (2), 449–460.

Wu, H., *et al.*, 2019. A CNN-SVM combined regression model for continuous knee angle estimation using mechanomyography signals. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 124–131.

Yamamoto, K., 2010. Road traffic noise prediction model "ASJ RTN-Model 2008": report of the research committee on road traffic noise. *Acoustical Science and Technology*, 31 (1), 2–55.

Ye, C., *et al.*, 2021. Urban function recognition by integrating social media and street-level imagery. *Environment and Planning B: Urban Analytics and City Science*, 48 (6), 1430–1444.

Yilmaz, G., and Hocanli, Y., 2006. Mapping of noise by using GIS in Şanliurfa. *Environmental Monitoring and Assessment*, 121 (1–3), 103–108.

Yin, L., *et al.*, 2015. Big data'for pedestrian volume: exploring the use of Google Street View images for pedestrian counts. *Applied Geography*, 63, 337–345.

Zambon, G., Benocci, R., and Brambilla, G., 2016. Statistical road classification applied to stratified spatial sampling of road traffic noise in urban areas. *International Journal of Environmental Research*, 10 (3), 411–420.

Zeng, L., *et al.*, 2018. A fast approach for large-scale sky view factor estimation using street view images. *Building and Environment*, 135, 74–84.

Zhang, F., *et al.*, 2019a. Social sensing from street-level imagery: a case study in learning spatio-temporal urban mobility patterns. *ISPRS Journal of Photogrammetry and Remote Sensing*, 153, 48–58.

Zhang, K., *et al.*, 2022. Using street view images to identify road noise barriers with ensemble classification model and geospatial analysis. *Sustainable Cities and Society*, 78, 103598.

Zhang, W., *et al.*, 2018. Using deep learning to identify utility poles with crossarms and estimate their locations from Google Street View images. *Sensors*, 18 (8), 2484.

Zhang, W., *et al.*, 2019b. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transportmetrica A: Transport Science*, 15 (2), 1688–1711.

Zhao, T., *et al.*, 2023. Sensing urban soundscapes from street view imagery. *Computers, Environment and Urban Systems*, 99, 101915.

Zhong, T., *et al.*, 2021. Assessment of solar photovoltaic potentials on urban noise barriers using street-view imagery. *Renewable Energy.*, 168, 181–194.

Zhou, B., *et al.*, 2016. Learning deep features for discriminative localization. *Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, NV: IEEE, 2921–2929.

Appendix

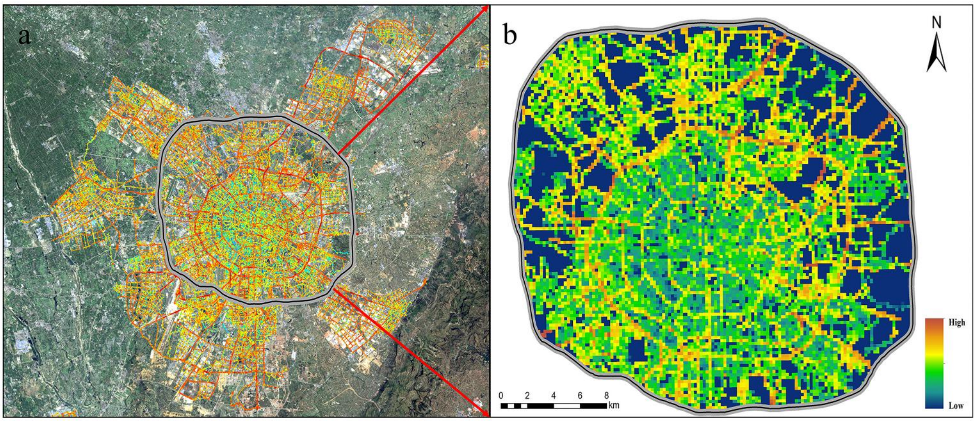

Figure A1. Road traffic noise estimation result of Chengdu (a) and grid map in the core area (b).

Figure A2. Difference map between the observed and the predicted road traffic noise within the outer ring freeway of Chengdu.

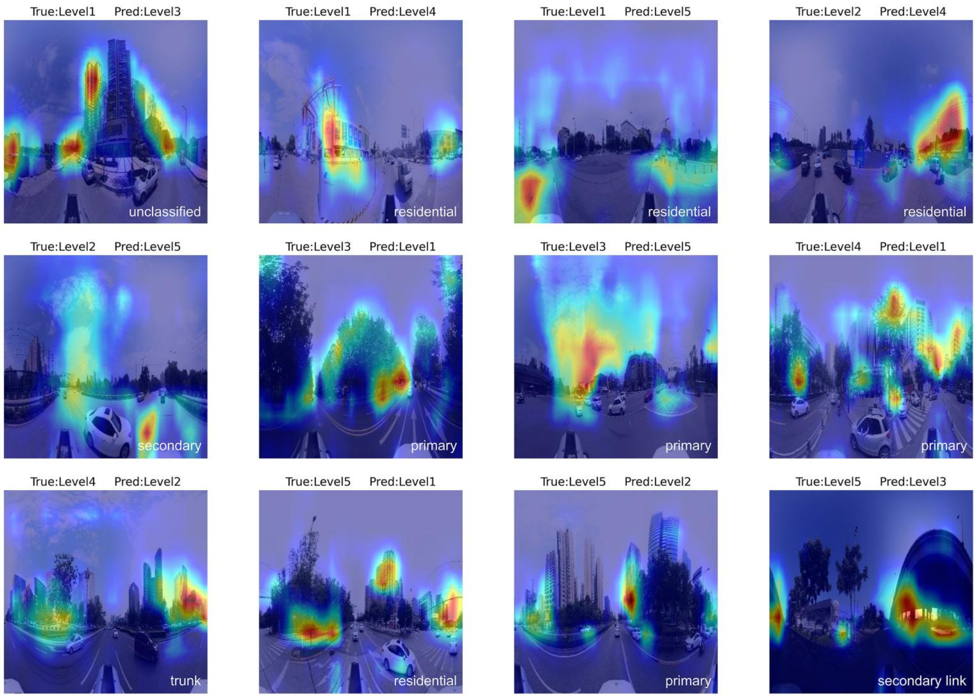

Figure A3. Misclassified samples of street view images and the corresponding road types with the Grad-CAM overlaid.