International Journal of Geographical Information Science

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

A topography-aware approach to the automatic generation of urban road networks

Zhou Fang, Jiaxin Qi, Lubin Fan, Jianqiang Huang, Ying Jin & Tianren Yang

To cite this article: Zhou Fang, Jiaxin Qi, Lubin Fan, Jianqiang Huang, Ying Jin & Tianren Yang (2022) A topography-aware approach to the automatic generation of urban road networks, International Journal of Geographical Information Science, 36:10, 2035-2059, DOI: 10.1080/13658816.2022.2072849

To link to this article: https://doi.org/10.1080/13658816.2022.2072849

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2022, VOL. 36, NO. 10, 2035–2059
https://doi.org/10.1080/13658816.2022.2072849

RESEARCH ARTICLE

A topography-aware approach to the automatic generation of urban road networks

Zhou Fang[a], Jiaxin Qi[b], Lubin Fan[c], Jianqiang Huang[c], Ying Jin[a] and Tianren Yang[d]

[a]Martin Centre for Architectural and Urban Studies, University of Cambridge, Cambridge, UK; [b]School of Computer Science & Engineering, Nanyang Technological University, Singapore; [c]DAMO Academy, Alibaba Group, Hangzhou, China; [d]Department of Urban Planning and Design, The University of Hong Kong, Hong Kong SAR, China

ABSTRACT
Existing deep-learning tools for road network generation have limited applications in flat urban areas due to their overreliance on the geometric and spatial configurations of street networks and inadequate considerations of topographic information. This paper proposes a new method of street network generation based on a generative adversarial network by designing a pre-positioned geo-extractor module and a geo-merging bypath. The two improvements employ the complementary use of geometric configurations and topographic features to automate street network generation in both flat and hilly urban areas. Our experiments demonstrate that the improved model yields a more realistic prediction of street configurations than conventional image inpainting techniques. The model's effectiveness is further enhanced when generating streets in hilly areas. Furthermore, the geo-extractor module provides insights from the computer vision perspective in recognizing when topographic information should be considered and which topographic information should receive more attention.

ARTICLE HISTORY
Received 11 August 2021
Accepted 28 April 2022

KEYWORDS
Street network generation; topographic information; computer vision; generative adversarial network; planning support systems

1. Introduction

The emerging deep learning-driven tools open a new window for geographers and planners to understand the urban fabric and provide an evidence base for road network design (Law *et al.* 2020, Shi *et al.* 2021). Despite a series of attempts to automate the process of street network generation (Hartmann *et al.* 2017, Kelvin and Anand 2020, Fang *et al.* 2021), the applications of existing deep learning-based approaches are limited in flat urban areas, and produce unreliable results in hilly regions (Fang *et al.* 2020b). The main reason lies in overreliance on the geometric and spatial configurations of streets and roads and has limited or zero consideration of topographic information in the automated generation process.

CONTACT Tianren Yang tianren@hku.hk
Supplemental data for this article can be accessed here.

This paper aims to fill the gap by proposing a topography-aware approach to the automatic generation of urban road networks. Building on the proven track record of generative adversarial network (GAN)-based methods of street network generation, this research presents an extended GAN-based model with a pre-positioned geo-extractor (GE) module and a geo-merging (GM) bypath, making complementary use of both the geometric configurations and topographic features of street networks to automate street network generation in both flat and hilly urban areas.

Note that street network generation in this study specifically refers to the prediction of missing segments in a street configuration within a pre-defined region. The generation capability is trained through mining the correlation between local conditions (i.e. topographic features and surrounding street network) and a spatial layout of streets based on real-world cases. Although the developed model acknowledges the existence of unexpected characteristics of the actual street network, it will focus on identifying the dominant influencing attributes and their relative importance. Therefore, the model is not designed to accurately reproduce all the road segments (as some of them are arbitrarily designed), though its overall performance is meaningful in reflecting how well the street configuration, in general, can be reproduced based on the given local conditions.

Four Italian cities (Florence, Perugia, Rome and Siena) were selected as case study areas to evaluate the performance of the proposed model. With a rich mixture of hilly and flat urban areas, Italy is a suitable place for exploring the roles of natural geography and human interventions in street network design. Specifically, we first develop a street network dataset, which includes approximately 56,000 street network samples extracted from the case study areas. For each sample, the two-dimensional rasterized street network data are enhanced with seven types of topographic information, including topographic elevation, slope, aspect of slope, and a set of hill shade data derived from four different azimuths of light (0, 90,180 and 270 degrees).

We then propose a three-stage GAN-based model of street network generation, including (1) a pre-positioned GE module for topographic information selection and weighting (stage 1) in line with the generation tasks in various topographic conditions, and (2) GM bypaths for vertices prediction and coarse street network generation (stage 2) and street network refinement (stage 3). The model design builds on the latest work by Fang *et al.* (2021), who first employed a GAN-based image inpainting technique for road network generation, and Yu *et al.* (2019), who developed a multi-stage generation framework for better model performance in image inpainting.

Specifically, we extend the standard encoder-decoder system from a state-of-the-art model for image inpainting (Yu *et al.* 2019) to consider additional features (i.e. local topographic information) beyond the surrounding geometric and spatial configurations alone. The improvement is achieved by attaching GM bypaths that learn the correlations between as-built street networks and local topographic features. Furthermore, the proposed three-stage model structure enables (1) the selection and weighting of topographic information, and (2) the use of vertices as intermediate guidance. The upgrade can address failure cases when the existing one-stage street generation models are applied to hilly areas (Fang *et al.* 2020b).

This research offers the following contributions. First, this study extends the capability of GAN-based street network generation methods to become aware of both the surrounding context and the local typography. The experiment's results demonstrate that the improved generation model can yield a more realistic prediction of street configuration whose effectiveness is further enhanced in hilly areas. Second, the weight variances of different typographic data are identified by the model, leading to context-specific planning and design implications for practitioners. Third, although the model itself is not developed for designing creative spatial proposals, it can learn from precedents such as organically developed road networks. Moreover, the model can follow planning guidance such as the junction locations to produce a series of alternative plans for comparative evaluations in practice.

The remainder of this paper is organized as follows. An overview of related work is offered in Section 2. Our urban street dataset and a brief description of the topographic data are introduced in Section 3. The proposed topography-informed model of urban street network generation is presented in Section 4, followed by details of experiments, results, and discussions in Sections 5 and 6. Section 7 concludes the paper.

2. Related and previous research

The development of digital assistance tools remains topical in planning fields where a series of alternative plans can be generated for comparative evaluations, leading to the delivery of optimal design and engineering solutions within a narrow time window (Coons 1964, Yang *et al.* 2019, Yang 2020). With the advances in information technology, procedural and learning-based digital assistance tools have been widely adopted in support of street network generation and visualization.

Procedural-based tools are used in design proposals based on manually designed rule sets. For example, Parish and Müller (2001) adopted a Lindenmayer system for expanding road networks with four different rule sets for automatic road network generation, namely, the basic rule, New York rule, Paris rule, and San Francisco rule. Users choose and apply the desired rule set by jointly considering topographic constraints and site context. The other prevalent rule sets in street network generation include the anisotropic shortest-path rule (Galin *et al.* 2010) and the direction field from virtual traffic simulation (Chen *et al.* 2008, Beneš *et al.* 2014). The rule sets used for street network generation can also be derived from statistics. For instance, Aliaga *et al.* (2008) employed preprocessed statistics regarding existing intersections (e.g. degrees and hierarchies) as rules for street generation that link the proposed road junctions.

In comparison, learning-based approaches draw lessons from real-world street network cases. Hartmann *et al.* (2017) made one of the first attempts to develop an automatic road generation tool, StreetGAN. The tool used a GAN system to synthesize street networks, maintaining the consistency of street networks' geometric and spatial configurations learned from the training data set. In addition, Kempinska and Murcio (2019) and Law and Neira (2019) trained variational autoencoders and convolutional-PCA using rasterized street network patches derived from OpenStreetMap (OSM), and attempted using low-dimensional vectors to control the street network generation.

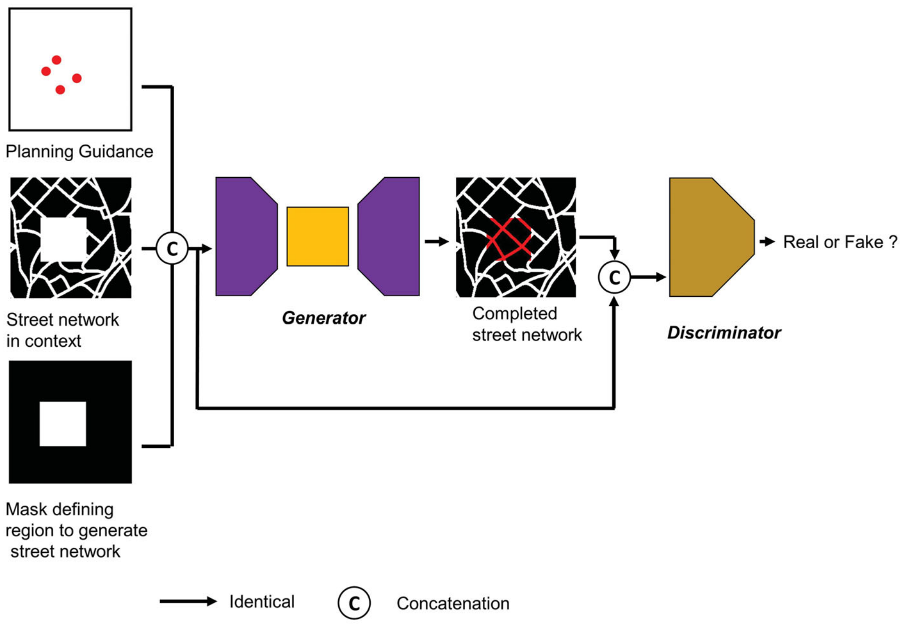

Figure 1. Conventional structure of GAN-based street network generation models.

Inspired by the image completion task in computer vision, Fang *et al.* (2020b) adapted one of the state-of-the-art GAN-based image inpainting models (Iizuka *et al.* 2017). The authors proposed a context-aware street generation module, DeepStreet, that can predict the future expansion patterns of street networks within the pre-defined region conditioned by surrounding street networks. Recently, Fang *et al.* (2021) incorporated planning guidance into the GAN-based generation process, including street junction locations and street-pattern-type annotations. This development turns the end-to-end generation system into a controllable process (Fang *et al.* 2021). The result leads to a more realistic prediction of street configurations and provides both professional and lay users with the opportunity to intuitively explore alternative street network designs for comparison and further evaluation. The conventional structure of GAN-based street network generation models is shown in Figure 1.

Despite increasing efforts to iteratively develop digital assistance tools for automatic street network generation, current tools fall short of emphasizing the role of topographic information in the generation process. A few procedural-based methods have offered topographically aware generation rules, such as the San Francisco rule in Parish and Müller (2001) and the anisotropic shortest path rule in Galin *et al.* (2010). However, the rules are general and must be uniformly adopted to the pre-defined area. Moreover, once the topographically aware generation rules are adopted, the geometric and spatial configurations of the street network in the context region cannot be considered simultaneously in the generation process. This limitation complicates the production of high-quality street network designs in the hilly-to-flat transition area.

Regarding learning-based approaches, GAN-based models limit their applications to flat-terrain urban areas, with failure cases occurring when the models are applied to hilly urban fringes (Fang *et al.* 2020b). The main reason lies in an overreliance on the geometric and spatial configurations of streets in the model training and inference process combined with limited or nonexistent consideration of topographic information. Most studies do not input any topographic information into the GAN-based system during the model training stage. As a result, the models are unable to establish links between topographic features and as-built street networks (Hartmann *et al.* 2017, Kempinska and Murcio 2019, Law and Neira 2019).

Among the limited attempts to consider topographic constraints, Fang *et al.* (2020b) incorporated topographic information into the generation process. Two additional channels, topographic elevation and aspect of slope, were attached to the training and testing data sample. However, failure cases demonstrated that the models could not effectively extract topographic features and were not equipped with the capability of using appropriate topographic information to guide street network generation in hilly areas. Data on topographic elevation and slope aspect alone are insufficient to guide street network generation; thus, other types of topographic information must be reviewed and incorporated.

This research aims to build on the proven track records of learning-based street network generation systems, proposing a topography-informed model to address the limitations and challenges identified in the literature. First, we include seven types of topographic information in the system and propose a GE module for the selection and weighting of this information. This procedure allows our model to recognize various types of topographic information and learn a soft-gate mechanism with values ranging from 0 to 1. The mechanism amplifies appropriate topographic information in hilly regions where such data are essential to guide street network generation. In addition, the GE module restrains topographic data in flat urban regions where street network patterns in context regions are essential for inspiring street network generation.

Furthermore, we improve a GAN-based model that takes preprocessed topographic data, street networks in the context region, and desired junction locations as inputs. A completed street network is then output. The existing end-to-end generation system is modified to a three-stage mechanism that first learns to predict the vertices of the street network based on preprocessed topographic information. The street network within the predefined region is then completed, using the predicted vertices as intermediate guidance.

Meanwhile, at the generation stages, we also upgrade the standard encoder-decoder system that is widely adopted in existing image inpainting models (Yu *et al.* 2018, 2019). This upgrade is accomplished by attaching an additional GM bypath to learn the correlations between as-built street networks and local topographic features. The improvement enables the model to jointly consider the geometric and spatial configurations of a non-local street network (through existing encoder and decoder paths designed for image inpainting), and local topographic features (through the additional GM bypath) during the generation process. With the above updates, our model can adaptively generate appropriate street networks within the predefined region. The generated street network can reasonably fit to the surrounding street network by jointly considering site context, planning guidance, and topographic constraints.

3. Urban road network dataset with enhanced topographic information

Two stages occur when preparing an urban road network dataset (Figure 2), namely, (1) base map preparation with enhanced topographic information and (2) sample collection. To evaluate the performance of the proposed model, we created an urban street network dataset enhanced with topographic information for four cities in central Italy: Florence, Perugia, Rome, and Siena. These four cities are located in the Apennines – a series of mountain ranges bordered by narrow, flat coastland, forming the backbone of peninsular Italy. Rome and Florence were selected as the examples of human-centric development in the flat part of central Italy, while Perugia and Siena were selected to represent such development under tight topographic constraints.

3.1. Base map preparation

In the first stage, we prepared base maps for the four cities in central Italy by creating multiple overlapping layers, or 'channels,' of an image to represent the comprehensive attributes associated with the road network design. These channels were divided into two specific types: topographic information and street-network-related data.

3.1.1. Topographic information (channels 1–7)

Topography is identified as one of the dominant factors influencing street network orientation, configuration, and entropy (Boeing 2019). We collected publicly accessible digital elevation model (DEM) data as a basis to derive topographic information channels. A comprehensive list of variables widely acknowledged in terrain analysis have

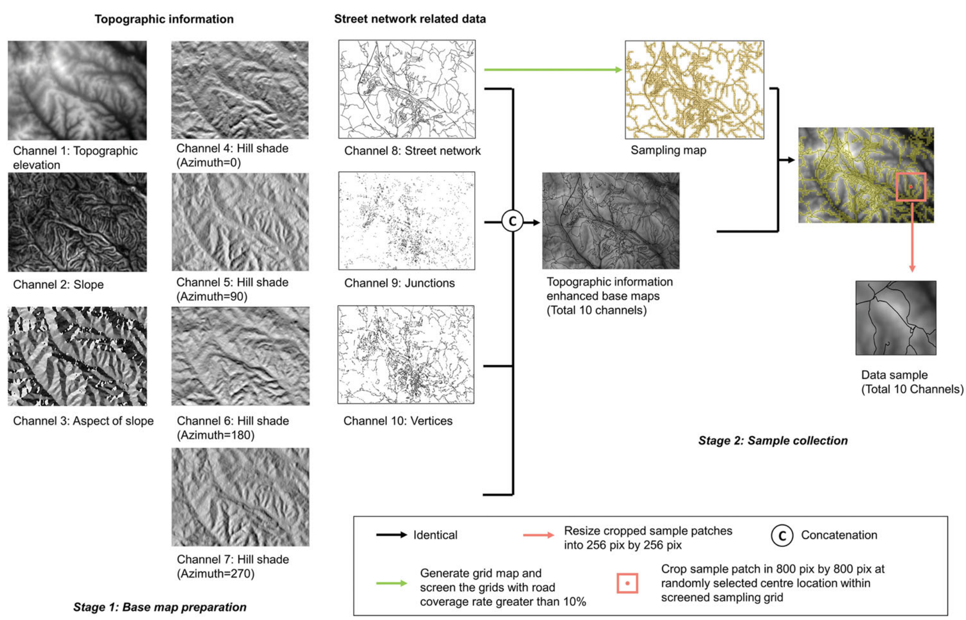

Figure 2. Overall process to set up the urban road network dataset enhanced with topographic information.

been adopted in this study (Olaya 2009), including topographic elevation (Channel 1), slope (Channel 2), aspect of slope (Channel 3), and four shade maps of hills (Channels 4–7). The resolution of the raw DEM data is 30 by 30 m. Based on spatial interpolation and aggregation, we prepared the DEM data at three different spatial resolutions to simultaneously train the model (see Section 4.2): (1) 2 by 2 m (a 256 by 256 pixel [pix] sample represents a 512 by 512 m area); (2) 8 by 8 m (a 64 by 64 pix sample represents a 512 by 512 m area); and (3) 32 by 32 m (a 16 by 16 pix sample represents a 512 by 512 m area).

3.1.2. Street-network-related data (channels 8–10)

An open-source vector representation of street networks, with street level attributes provided by OSM, was used to retrieve street-network-related information. Street networks (including motorway, trunk, primary, secondary, tertiary, and residential following the OSM definitions) were rasterized to form a road network map (Channel 8). The width of streets was consistently set as 10 pix (representing 10 m) to guarantee that all road segments could be fully represented when the images are compressed in the training stage. In addition, all the junction points (with the degree of connectivity being one, three, four and five) and vertices in the road networks are rasterized into heat maps (Channels 9 and 10) by applying 2 D Gaussian with standard deviation of 10 pix centered on their respective locations.

3.2. Data sample collection

In the data sample collection stage, we first identified the areas covered by a road network by defining a road coverage rate greater than 10% within a 100 by 100 m grid. Among the road-covered grids, we randomly assigned points as center locations to crop 512 by 512-m-multichannel-data samples, which are then compressed into 256 by 256 pix patches to save computational costs. According to a recent urban fabric classification study by Fang *et al.* (2020a), a 512 by 512 m data sample is the most appropriate spatial unit for the application of deep learning-based computer vision techniques compared to other patch sizes of 128, 256, and 1,024 m. The random extraction of data patches yields approximately 56,000 street network samples for the four case study areas (Figure 3). More detailed statistics regarding the four case study areas are summarized in Supplemental Material A.

4. Developing a topography-aware GAN model

4.1. Overall structure

We built the proposed topography-informed street network generation model based on a state-of-the-art model for image inpainting (Yu *et al.* 2019). Following the general principles of GAN, a generation system and a discrimination system are jointly and iteratively trained. The generation system produces street networks within a defined region and the discrimination system evaluates the generalized street network until it is hard to distinguish the outputs from the ground truth. The generation system is a

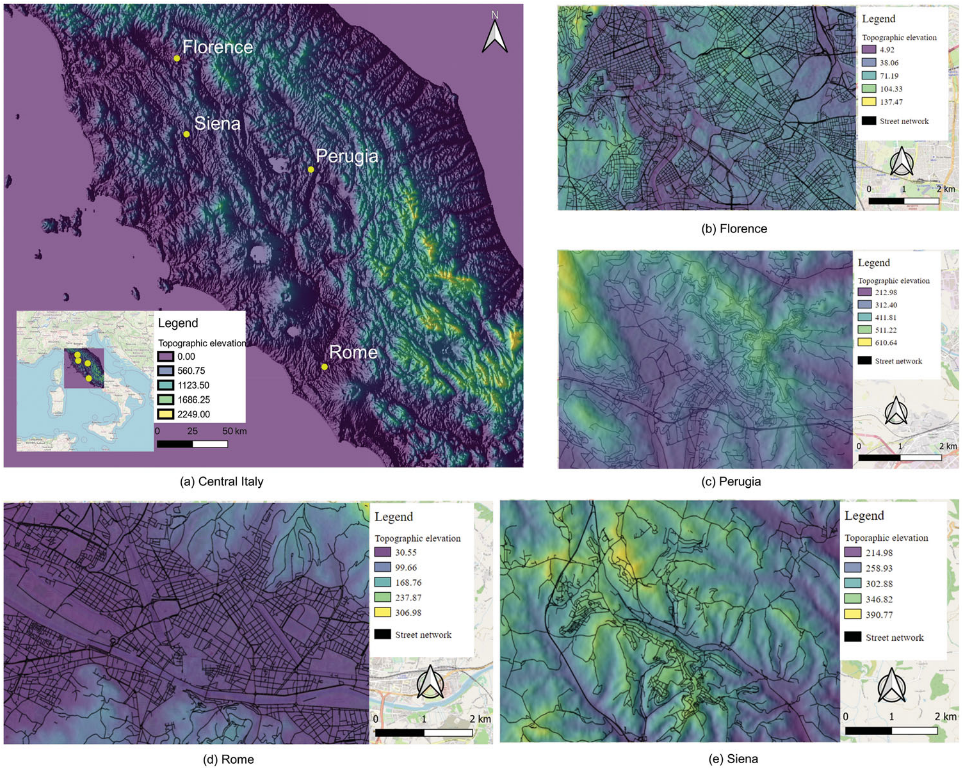

Figure 3. Topographic maps with highlighted sampling areas for the four case study cities.

cascade of three stages: (1) topographic information preprocessing, (2) vertices prediction and coarse street network generation, and (3) street network refinement.

As shown in Figure 4, the inputs of the proposed generation system include contextual street network (In_{st}) defined by mask channels (In_{mask}), junction guidance (In_{guid}), and topographic information (Raw_{topo}). The model output is the completed street network (Out_{sts2}). A detailed description of model variables is summarized in Supplemental Material B.

Prior to the three-stage model training, model input and ground truth pairs (stage 0) are prepared given data samples from the prepared training dataset (see Section 3). We randomly generate masks, In_{mask}, following the algorithm described by Fang *et al.* (2020b) to derive input street network patches $In_{st} = x_{st} \odot In_{mask}$. Specifically, a center location of the missing area is randomly assigned to generate a rectangular mask (with the width and height being a random number from 48 to 64 pix) within a street network sample of 256 by 256 pix. Supplemental Material C provides further details on the mask-generation process.

To foster a sharp and realistic street network generation, a Spectral-Normalized Markovian Discriminator (SN-PatchGAN) (Yu *et al.* 2019), widely adopted in other image inpainting tasks (Jo and Park 2019), is used in our model. A convolutional network with spectral normalization (Miyato *et al.* 2018) is used as the feature extractor,

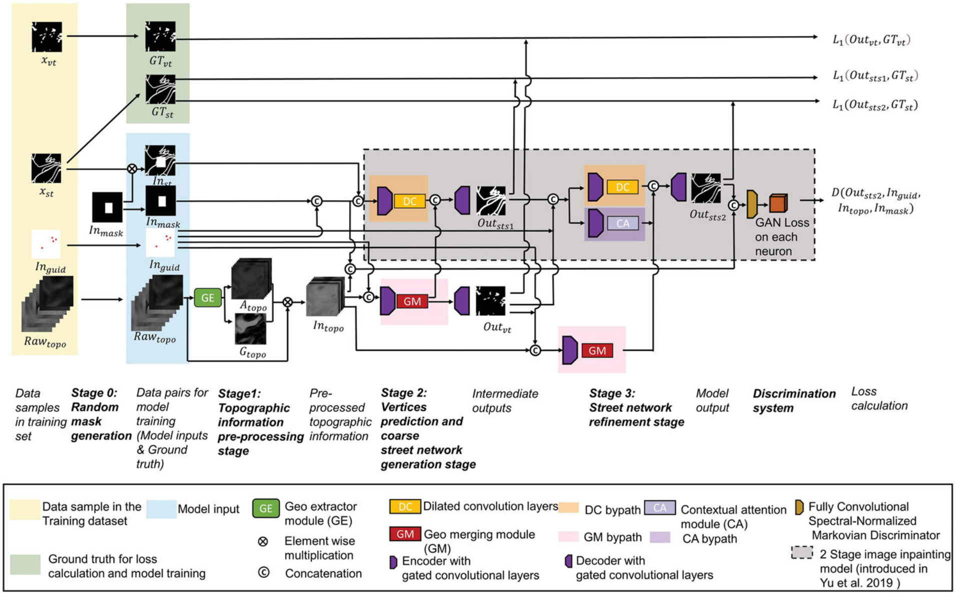

Figure 4. Overall framework of proposed street network generation system.

where the input consists of a street network patch ($Output_{sts2}$), a mask (In_{mask}), junction guidance (In_{guid}), and preprocessed topographic information (In_{topo}). The output is a three-dimensional feature cube in shape $R^{h \times w \times c}$ (h,w,c representing the height, width and number of channels). We then apply GANs to each element in the feature cube, forming a set of 'true or false' predictions on different spatial locations of the input street network patches (see Section 4.4 for details).

4.2. Stage 1: a geo-extractor (GE) module for topographic information selection and weighting

In stage 1, the GE module is designed to learn dynamic topographic information selection and weighting mechanisms at different spatial locations. In hilly areas, topographic information is deemed more essential to guide street network generation, whereas street network patterns are deemed more essential to inspire street network generation in flat regions. Given the seven topographic information channels of sample Raw_{topo}, we aim to first predict a gate channel, G_{topo}, defining the spatial distribution of hilly areas and flat regions. Second, we seek to predict a set of heat maps, A_{topo}, for the seven topographic information channels, highlighting topographic features that can be used to guide street network generation. G_{topo} is a channel gate map sharing the same shape and resolution as the topographic information channels with decimals ranging from 0 to 1, describing how much of the topographic information should be let through. A value of 0 means no thoroughfare, while a value of 1 means full transmission with no losses. A_{topo} consists of seven heat maps for the seven input topographic information channels, containing decimals ranging from 0 to 1 representing the intensity of attention.

In the GE module, we adopt parallel multi-resolution subnetworks with multi-scale fusion mechanisms for subnetwork connection and information exchange. The module takes preprocessed topographic information patches in shapes 256 by 256 pix, 64 by 64 pix, and 16 by 16pix as inputs for the three high-to-low resolution subnetworks. Lower resolution subnetworks are gradually added one by one and stage by stage, from 4 by 4 pix to 1 by 1 pix. This structure enables the module to maintain both high-to-low resolution spatial representations of the topographic information and learn channel-wise attention among the seven types of topographic data. The detailed structure of the GE module is shown in Supplemental Material D.

4.3. Stages 2 and 3: Geo-merging (GM) bypaths for vertices prediction and coarse street network generation as well as street network refinement

Backbones with dilated convolutional layers and contextual attention modules adopted by Yu *et al.* (2019) are among the state-of-the-art research for image inpainting tasks. These backbones are designed to support the models in learning non-local correlations among pixels in the context region as well as the missing region throughout the training stage. The process expands the reception field to 'see' a larger area of the input image when computing each output pixel within the predefined infill region. This strategy works well for street network generation on flat terrain where the street network designs are always inspired by surrounding street network patterns, aiming to maintain the consistency of local spatial configurations. However, a street network in a hilly urban fringe area is always sparse, thus lacking enough surrounding street network patterns that can be borrowed to infill the missing area. In practice, designers usually regard local topographic conditions as primary constraints when designing street networks in hilly regions, subject to the consideration of infrastructure elements (e.g. drainage), engineering feasibility (e.g. soil type and vegetation cover), and planning regulations (e.g. developable areas). This practice results in real-world street networks in hilly regions showing a strong association with local topographical features (Figure 5).

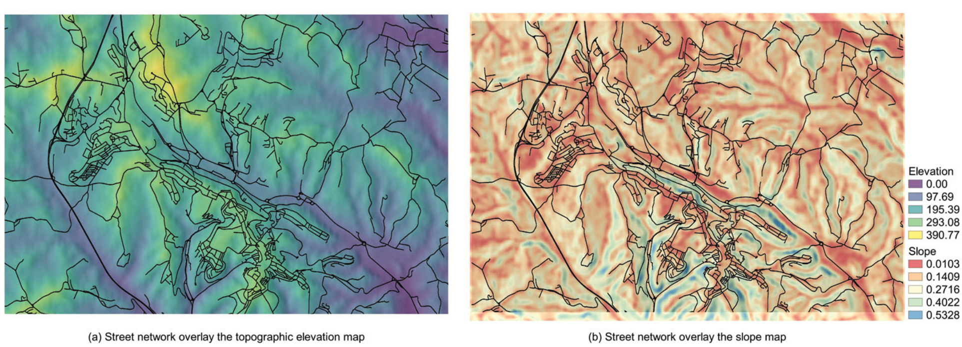

Figure 5. Topographic elevation map and slope map for Siena.

To address the limitations of GAN-based models in incorporating topographic information, we append a GM bypath to existing models. Through model training, the correlations among street networks and local topographic features can be mined.

In stage 2—prediction of vertices and coarse street-network generation—two parallel encoder–decoder structures (paths) are introduced. The GM bypath takes preprocessed topographic information (In_{topo}) and junction guidance (In_{guid}) as inputs and predicts a vertices heatmap (Out_{vt}). Meanwhile, the dilated convolutional layers bypath takes the masked street network (In_{st}), junction guidance (In_{guid}), and randomly generated mask (In_{mask}) as inputs. Features aggregated with encoded topographic information (via GM bypath) are output and transferred to the decoder for coarse street network prediction (Out_{sts1}).

Similarly, stage 3, the street network refinement stage, consists of three parallel paths: a dilated convolution layer bypath, a contextual attention bypath, and a GM bypath in encoder–decoder structures. The network structures of the dilated convolution layer bypath and the contextual attention bypath are identical to those used in Yu *et al.* (2019). These structures employ the predicted coarse street network (Out_{sts1}), junction guidance (In_{guid}), generated mask (In_{mask}), and predicted vertices heatmaps (Out_{vt}) as inputs. Output features from the three paths are aggregated and fed into a single decoder to obtain the final completed street network (Out_{sts2}). More details regarding GM bypaths can be found in Supplemental Material D.

4.4. Loss functions

Two types of loss, namely reconstruction and adversarial, are jointly minimized and maximized respectively to train the proposed deep neural networks by iteratively updating the parameters embedded in the generation and discrimination systems. The mixture of the two loss functions allows for the stable training of a high-performance network model and has been widely adopted for image completion. The trained generation system is then used for performance evaluation and further inference.

Reconstruction loss (L_{recon}) is designed to quantitatively evaluate the model outputs by measuring the average pixel-wise distance between the outputs and the ground truths. At the model training stage, the model is calibrated towards generating outputs that are identical to ground truths by minimizing the reconstruction loss. In this work, the total reconstruction loss consists of three components: loss for predictions of the vertices heatmap, coarse street networks, and the refined street network. The final reconstruction loss function is defined as:

$$L_{recon} = \alpha \cdot L_1(Out_{vt},\ GT_{vt}) + \beta \cdot L_1(Out_{sts1}, GT_{st}) + \gamma \cdot L_1(Out_{sts2}, GT_{st}) \tag{1}$$

where α, β and γ are hyperparameters adopted to weight the three reconstruction losses. $L_1(x, y)$ represents the mean absolute error (L_1 loss), which is widely used in the machine learning field to quantify the pixel-wise similarity between the generated data patches (Out_{vt} – vertices heatmap; Out_{sts1} – coarse street network; and Out_{sts2} – refined street network) and the ground truth patches (GT_{vt} and GT_{st}).

Adversarial loss (L_{adv}) is considered and included in the loss functions of most image-completion algorithms. By considering adversarial loss, the standard minimization process of reconstruction loss can be turned into a matter of min-max optimization in which the discriminator is jointly updated with the generation network at each training iteration. In this work, the hinge loss function is adopted to determine whether the input street network patch is real or fake. The adversarial losses for training the discrimination and generation systems are, respectively:

$$L_{adv} = 0.5 \times (-E[min(0, \; -1 + D(GT_{st}, \; In_{guid}, \; In_{topo}, \; In_{mask}))] \\ - E[min(0, \; -1 - D(Out_{sts2}, \; In_{guid}, \; In_{topo}, \; In_{mask}))]) \tag{2}$$

$$L_{gen} = \lambda \; \times \; L_{recon} + \; \mu \times (-E[D(Out_{sts2}, \; In_{guid}, \; In_{topo}, \; In_{mask})]) \tag{3}$$

where λ and μ are hyperparameters adopted to calculate the generation loss L_{gen}. $D(x, c_1, c_2, c_3)$ represents the calculation process in the discriminator, namely the SN-PatchGAN in this study. The discriminator takes either generated (Out_{sts2}) or ground truth (GT_{st}) street network patches as inputs, considering their attributes related to junction locations (In_{guid}), preprocessed topographic information (In_{topo}), and randomly generated masks (In_{mask}). The discriminator outputs values ranging from 0 (i.e. generated) to 1 (i.e. ground truth) for the input street network patches.

4.5. Model training

The concatenation of In_{st}, In_{mask}, In_{guid}, and Raw_{topo} forms the model input, while GT_{vt} and GT_{st} represent the ground truth for the loss calculation. The three-stage generation system G takes the model input and outputs the predicted street network Out_{sts2}. The discrimination system D takes the output from the generation system (Out_{sts2}), its ground truth pair (GT_{st}), and its generation conditions as the inputs. The system outputs a set of 'true or false' predictions on various spatial locations of the predicted street network patches Out_{sts2}. The training procedure is shown in Algorithm 1 and Figure 4. The model is converged when (1) the adversarial loss for D on the training dataset is maintained at a stable level, and (2) the reconstruction loss for G on the test dataset does not decrease with the number of iterations.

Algorithm 1. Training the proposed framework

1: **while** G has not converged **do**
2: Sample batch street network sample, x_{st}, In_{guid}, x_{vt} and Raw_{topo} from training data
3: Generate random masks In_{mask} for x_{st}
4: Construct input street network $In_{st} \leftarrow x_{st} \odot In_{mask}$
5: Get gate map and attention maps $G_{topo}, A_{topo} \leftarrow GE(Raw_{topo})$
6: Get preprocessed topographic information
$In_{topo} \leftarrow Raw_{topo} \odot A_{topo} \odot G_{topo}$;
7: Get predictions from the generator (G) for the two generation stages
$Out_{sts1}, Out_{vt}, Out_{sts2} \leftarrow G(In_{st}, In_{mask}, In_{guid}, In_{topo})$
8: Calculate adversarial loss for discriminator (D) training
$L_{adv} = 0.5 \times \left(-E[min(0, \; -1 + D(GT_{st}, In_{guid}, In_{topo}, In_{mask}))] - E[min(0, \; -1 - D(Out_{sts2}, In_{guid}, In_{topo}, In_{mask}))]\right)$
9: Update discriminator (D) with L_{adv}
10: Calculate Reconstruction loss
$L_{recon} \leftarrow \alpha \cdot L_1(Out_{vt}, GT_{vt}) + \beta \cdot L_1(Out_{sts1}, GT_{st}) + \gamma \cdot L_1(Out_{sts2}, GT_{st})$;
11: Calculate generation loss for generator (G) training
$L_{gen} \leftarrow \lambda * L_{recon} + \mu * D(Out_{sts2}, In_{guid}, In_{topo}, In_{mask})$
12: Update generator (G) with L_{gen};
13: **end while**

5. Experimental design and results

5.1. Experimental design

Three models are designed for comparing the performance of the conventional and proposed street network generation tools. All the models are tested using the same topographic enhanced street network dataset as introduced in Section 3. Note that the samples extracted from Florence and Perugia are used to train the models, while samples extracted from Rome and Siena are used for validation and testing.

Model 1 is built following the one-stage model structure introduced by Fang *et al.* (2021) and focuses on the fundamental context and junction nodes as planning guidance[1] for generating street networks. We also incorporate topographic information into Model 1 by simply attaching seven channels of topographic information to the input samples. This aims to capture the correlation between the topographic information and the existing street network through model training and use the topographic information to guide street network generation in model testing. Model 2 improves Model 1 by adopting a two-stage model structure that was first proposed by Yu *et al.* (2019), which achieved state-of-the-art performance in classic image inpainting tasks. The proposed three-stage model (Model 3) extends Model 2 by pre-positioning a pre-processing module for topographic information, GE, and appending a bypath, GM, for local topographical feature extraction and analysis. Model 3 further predicts vertices for street networks as indeterminate output, in which we design a loss component to guide the model training (Table 1).

Table 1. Alternative models for training and testing.

	Model 1	Model 2	Model 3
Model structure			
Generation structure	One stage	Two stages	Three stages
Prepositioned topographic information preprocessing module	–	–	✓
Input: Fundamental context and planning guidance			
Surrounding street networks	✓	✓	✓
Junction nodes	✓	✓	✓
Input: Topographic information			
Topographic elevation	✓	✓	✓
Slope	✓	✓	✓
Aspect of slope	✓	✓	✓
Hill shade maps with four different azimuths	✓	✓	✓
Output: Intermediate result			
Coarse street networks	–	✓	✓
Vertices	–	–	✓
Output: Final result			
Generated street networks	✓	✓	✓

5.2. Settings for model training and testing

Our models for all designed experiments were trained with Python v3.6.9, PyTorch v1.3.0, CUDA v10.0 and CUDNN v10.0 on 8 $\times$ NVIDIA Tesla P100 GPU for 50 epochs on each experiment. Each training epoch took 741 iterations with a batch size of 64. Our evaluation used 9,033 samples, and the test was conducted on the same device with a batch size of 64.

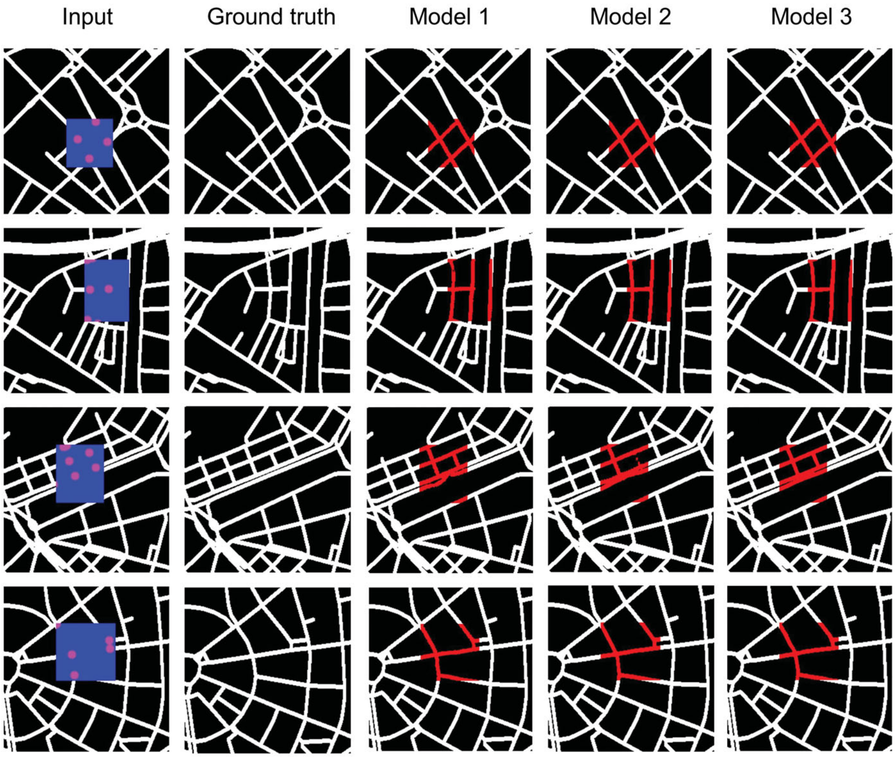

Figure 6. Street network generation in a flat urban area using Model 1, Model 2, and Model 3.

5.3. Testing results

5.3.1. Qualitative comparisons

Figure 6 compares the input data, the ground truth and model outputs for randomly selected flat-terrain urban areas in an intuitive manner[2]. With inputting ground truth junction locations and surrounding street networks, all three models successfully connected the junction nodes using straight-line segments and adapted the generated network to the surrounding network. All results are consistent with the vicinity with appropriate connections and coherent hierarchies. The performance of all three models is not affected by the topographic inputs (i.e. the seven types of topographic information).

The predictive performance of the three models differs when these are applied to hilly urban areas. As shown in Figure 7, although Model 1 and Model 2 successfully recognize the dead ends of the roads, linkages have not been fully established. In contrast, Model 3 can perform well in street network generation using either straight or curved lines (roads), with reasonable connectivity to the surrounding street networks.

The results above demonstrate that simply attaching additional topographic information to the input samples (Model 1) or applying state-of-the-art image inpainting techniques (Model 2) cannot yield reliable generation results in hilly urban areas. The sparse distribution of street networks in such areas makes it difficult for the conventional models (Models 1 and 2) to borrow information from surrounding areas for

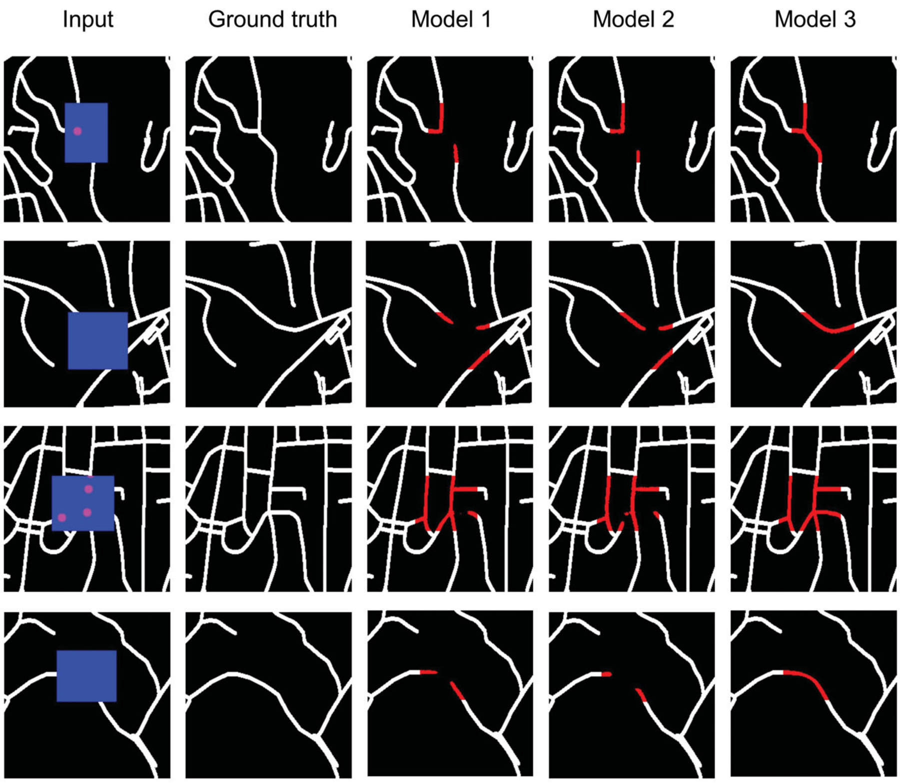

Figure 7. Street network generation in a hilly urban fringe area using Model 1, Model 2, and Model 3.

street network generation tasks. The proposed model (Model 3) equipped with the GE module and GM bypath proves successful in learning the correlations between local topographic features and street networks (Figure 8). The model can correctly identify the extracted local topographic features (through the GE module) and use them to guide the generation of winding mountain roads.

5.3.2. Quantitative evaluation

Similar to other image generation and image inpainting tasks, learning-based street network generation lacks good quantitative evaluation metrics, due to there being many possible solutions that differ from the original street network (Yu *et al.* 2018, 2019). In this work, we provide junction nodes and seven types of topographic information that serve as constraints in street network generation. As a result, the predicted street network is expected to be identical to the original one. Thus, evaluation metrics in terms of reconstruction errors are adopted in this work to assess the pixel-wise difference between predicted street networks and ground truths. Specifically, we report our evaluation in terms of mean L_1 error and mean L_2 error, in other words, mean absolute error (MAE) and mean squared error (MSE) respectively, for the three models shown in Table 2. The multiple-stage models (Model 2 and Model 3) perform

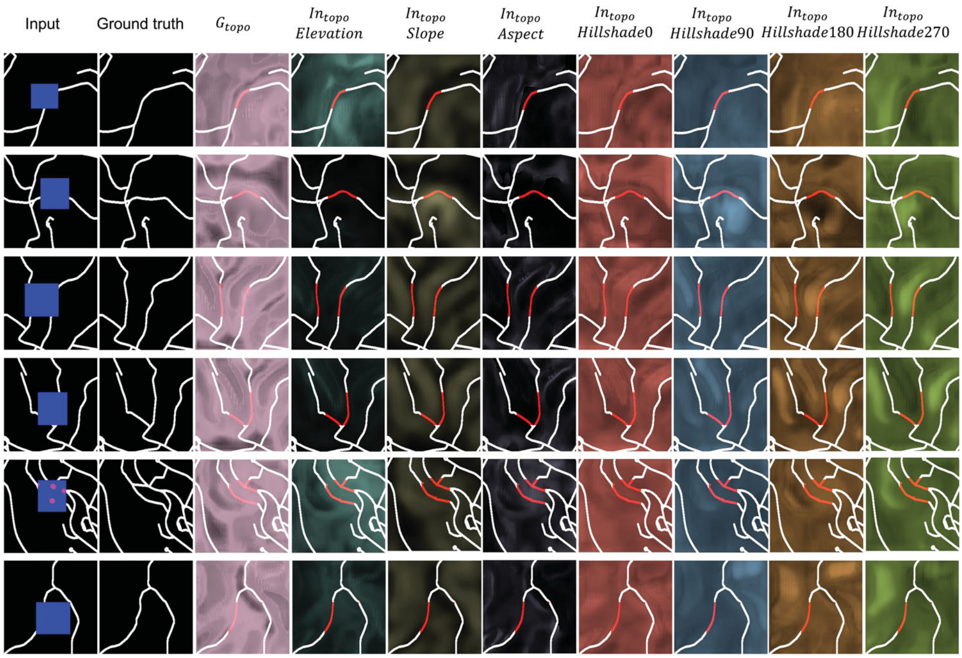

Figure 8. Street network generation results from Model 3 against preprocessed topographic information (In_{topo}) and derived gate channel (G_{topo}).

Table 2. Results of mean absolute error (MAE) and mean squared error (MSE) on the entire testing dataset.

Model	MAE	MSE
Model 1	0.1248	0.2453
Model 2	0.1236	0.2447
Model 3	0.1199	0.2345

better than Model 1 (one-stage model). The full model, Model 3, achieves the best performance of the three in terms of MAE and MSE.

Moreover, we calculate metric choice (MC) and metric integration (MI) to measure the differences in overall connectivity and accessibility between the generated street networks and the ground truth. Concretely, we first convert the output image representations of street networks to graphs, where the street network junction nodes and road segments are represented by graph vertices and graph edges, respectively. Then, we calculate MC and MI for all the road segments, and further calculate the length-weighted metric choice (LWMC), length-weighted metric integration (LWMI), absolute percentage errors for LWMC (APEMC), and LWMI (APEMI) for the generated street network following the equations below:

$$LWMC_k = \frac{\sum_{i=0}^{n} l_{k,i} \times MC_{k,i}}{\sum_{i=0}^{n} l_{k,i}} \tag{4}$$

$$LWMI_k = \frac{\sum_{i=0}^{n} l_{k,i} \times MI_{k,i}}{\sum_{i=0}^{n} l_{k,i}} \tag{5}$$

$$APEMC_k = \frac{\left|L\hat{W}MC_k - LWMC_k\right|}{LWMC_k} \tag{6}$$

$$APEMI_k = \frac{\left|L\hat{W}MI_k - LWMI_k\right|}{LWMI_k} \tag{7}$$

where, $l_{k,i}$ is the length of line segment l in sample k. $MC_{k,i}$ and $MI_{k,i}$ are the metric choice and metric integration for road segment l in sample k.

Examples of the image-to-graph conversion outputs (including the associated LWMC, LWMI, APEMC, and APEMI) from Model 1 and Model 3 are shown in Figure 9. Figure 10 further summarizes the APEMC and APEMI results across the entire samples in the test dataset. The street networks derived from Model 3 are found to be more likely to share a higher similarity in connectivity and accessibility (i.e. lower APEMC and APEMI) with the ground truths compared to those from Model 1.

Furthermore, Model 1 is associated with a more significant difference in connectivity and accessibility among generated street networks and the ground truths when applied to hilly urban areas. Compared to the case of Rome (i.e. a flat urban area), the application of Model 1 in Siena (a hilly urban region) yielded 17.02% (=21.93% − 4.91%) and 8.00% (=10.30% − 2.30%) more generational outputs with APEMC and AMEMI greater than 100%. Model 3 effectively narrows the performance differences between flat and hilly areas to 3.56% for APEMC and 0.38% for AMEMI.

6. Weighting the importance of topography in road network generation

To further demonstrate the working mechanism of the proposed GE module and justify the effectiveness of this module, we analyzed the performance of outputs from the GE, namely gate channels (G_{topo}) and heat map sets (A_{topo}), to interpret the learnt dynamic topographic information selection and weighting mechanism. To achieve this, we inferred the trained Model 3 with GE module (introduced in Section 5) on a newly created dataset for GE analysis (named 'GEanalysis set'). We followed the sampling process described in Section 3 but covered the whole rectangular sampling areas for Rome and Siena[3]. The analysis can also reveal the relative importance of the seven types of topography-related inputs.

6.1. Self-adaptive systems

The proposed GE module turns the original street network generation model for flat areas into a self-adaptive system, where the GE module outputs a gate channel, G_{topo}, for each input data sample. This controls the percentage of preprocessed topographic information passed to the following generation model to inform the street network prediction in both hilly and flat regions. The higher topographic information through rate represents greater need to incorporate topographic information in street network prediction, while the lower topographic through rate represents fewer topographic constraints in the design of street networks. The average value of the gate channel for each input data sample, $InRatio_{topo}$, also serves as an index for quantifying the

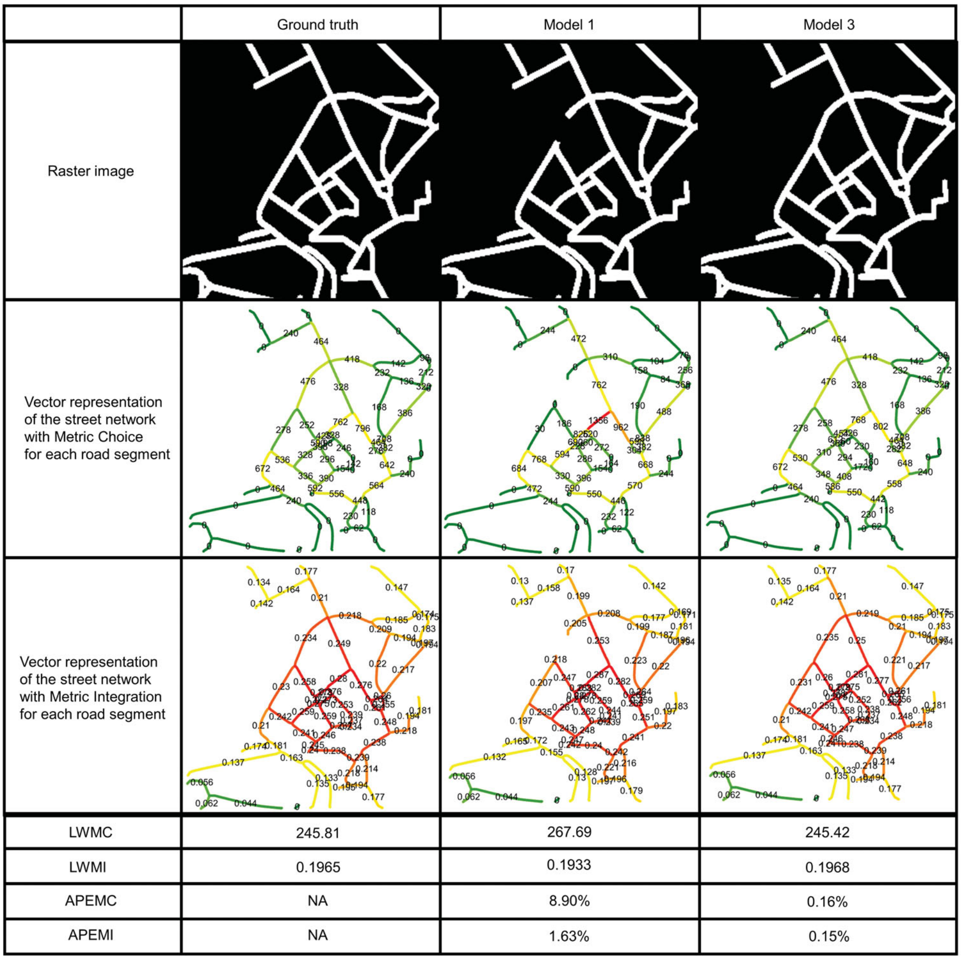

Figure 9. Examples demonstrating the process for calculating APEMC and APEMI for the output street network image.

topographic constraints for street network planning and informing practitioners when the topographic information should be considered.

To demonstrate the effectiveness of the self-adaptive system, we calculate the average topographic information through rates ($InRatio_{topo}$) for the samples in the GEanalysis set using the trained Model 3 with GE module. As shown in Figure 11, the road networks are formed in regular grid patterns and the majority of street junctions are connected using straight road segments within the area of lower $InRatio_{topo}$, while the road networks become sparser and the road becomes winding with an increasing $InRatio_{topo}$.

6.2. Weighting the seven types of topographic information

To explore the dynamic weighting mechanism of topographic inputs in the GE module (in Model 3), we calculate the channel-wise average weights for different types of topographic information for the samples in the GEanalysis set. W_k represents how the

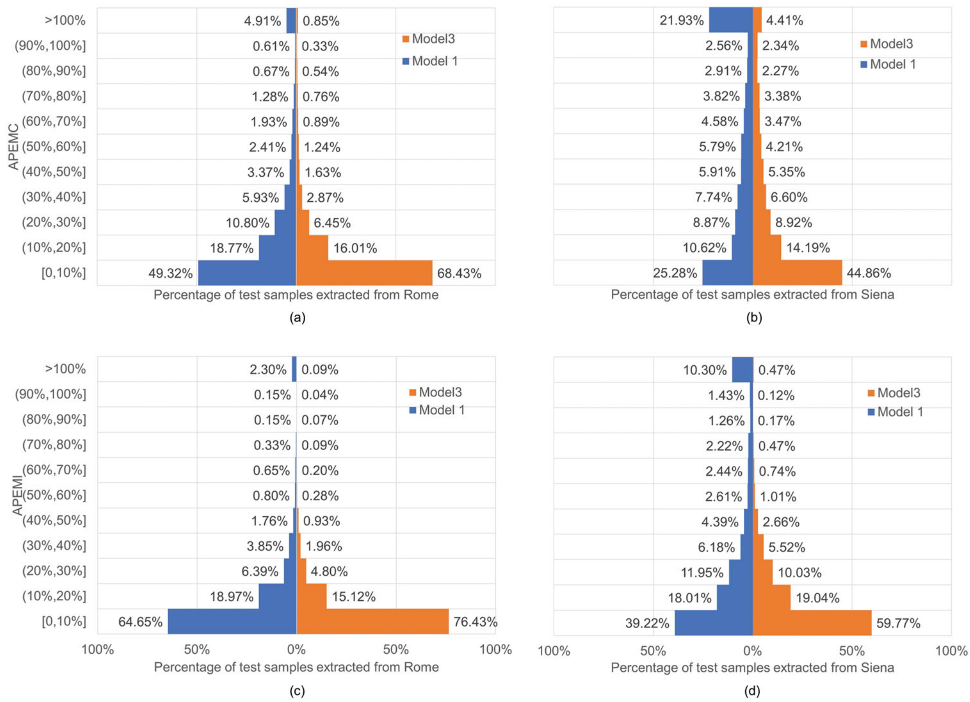

Figure 10. Violin frequency plots of APEMC and APEMI distributions for street network images outputted from Model 1 and Model 3, with input samples extracted from Rome and Siena.

GE module weighs different types of topographic information to predict realistic street networks in both flat and hilly regions:

$$W_k = \frac{\sum_{i=0}^{w}\sum_{j=0}^{h}(A_{topo_{k,i,j}} \times G_{topo_{i,j}})}{w \times h} \tag{8}$$

where W_k is the channel-wise average weight for different types of topographic information (k = 1 represents elevation, k = 2 represents slope, k = 3 represents aspect of slope, k from 4 to 7 represents the hill shade map with an azimuth of 0, 90, 180, 270 degrees, respectively); w is the width of the input data sample (256 pix in our case); h is the height of the input data sample (256 pix in our case); $A_{topo_{k,i,j}}$ is the output attention intensity for pixel at location i, j in channel k; $G_{topo_{i,j}}$ is the output topographic information through rate for pixel at location i, j. A higher W_k indicates the higher importance of topographic information k in street network generation tasks, while W_k=100% shows when the GE module takes the entirety of the raw topographic input k and transfers it to the generation model.

As shown in Figure 12, the GE module picks topographic information in different percentages. Among the seven types of topography-related inputs, the aspect of the slope is found to be weighed the least by the GE module: approximately 5% of the information is adopted by the GE module for the generation tasks. Compared with the raw topographic data, the GE module accepts 10%–30% of the information in all

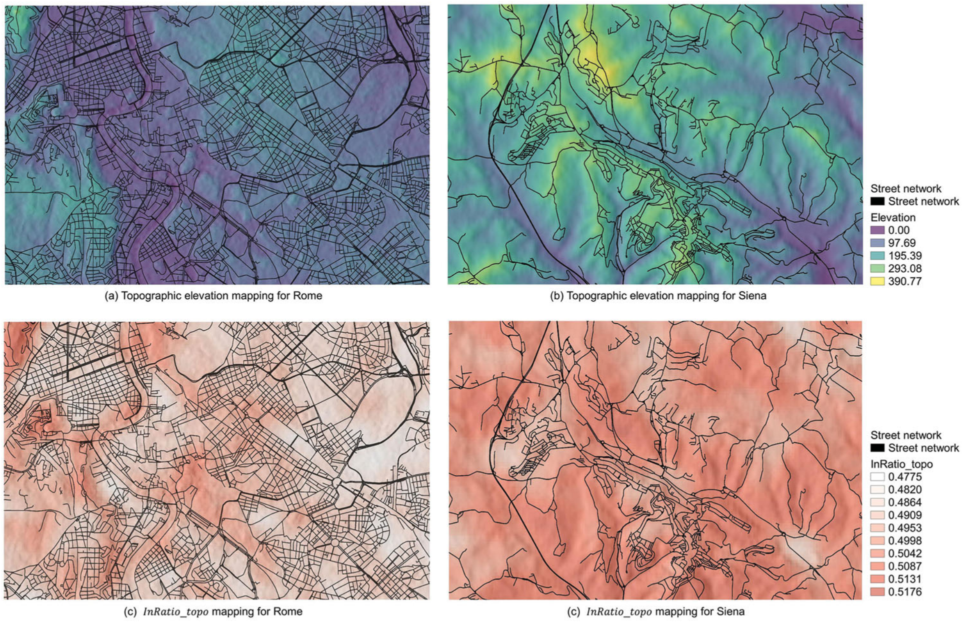

Figure 11. Topographic elevation and average topographic information through rate ($InRatio_{topo}$).

the other topographic channels, with percentages of the slope being weighted higher (24%–29%).

Moreover, the weights of the information on elevation and slope increase with their numerical values. Positive linear associations between the weights for the elevation and slope channels and correspondence input raw data are shown in Figure 12(a,b). Topographical information becomes more critical in guiding the street network design where there is a higher elevation and steeper slope. Subsequently, the GE module enhances the features embedded in elevation and slope channels by amplifying the raw data with higher intensity (i.e. higher numerical values times larger weights) while restraining raw data with lower intensity. A slope percent of 6% leads to a topographic weight of 24% in the GE module, but this weight will increase to over 27% for a slope percent greater than 16%. Linear regression also suggests that a slope percent of 9% serves as a threshold in street network generation. The importance of slope data is found to be greatly amplified in the GE module when the average slope percent is higher than 9% (i.e. 5.14 degrees).

7. Conclusions

In the era of information and communication technologies, deep learning-based models are increasingly used to inform spatial planning, such as street network generation (Law *et al.* 2020, Shi *et al.* 2021). In this research, we have proposed a new method for the automated generation of road networks that are both context aware and topography informed. Specifically, the method can effectively incorporate various types of topographic information into a street generation model based on a generative

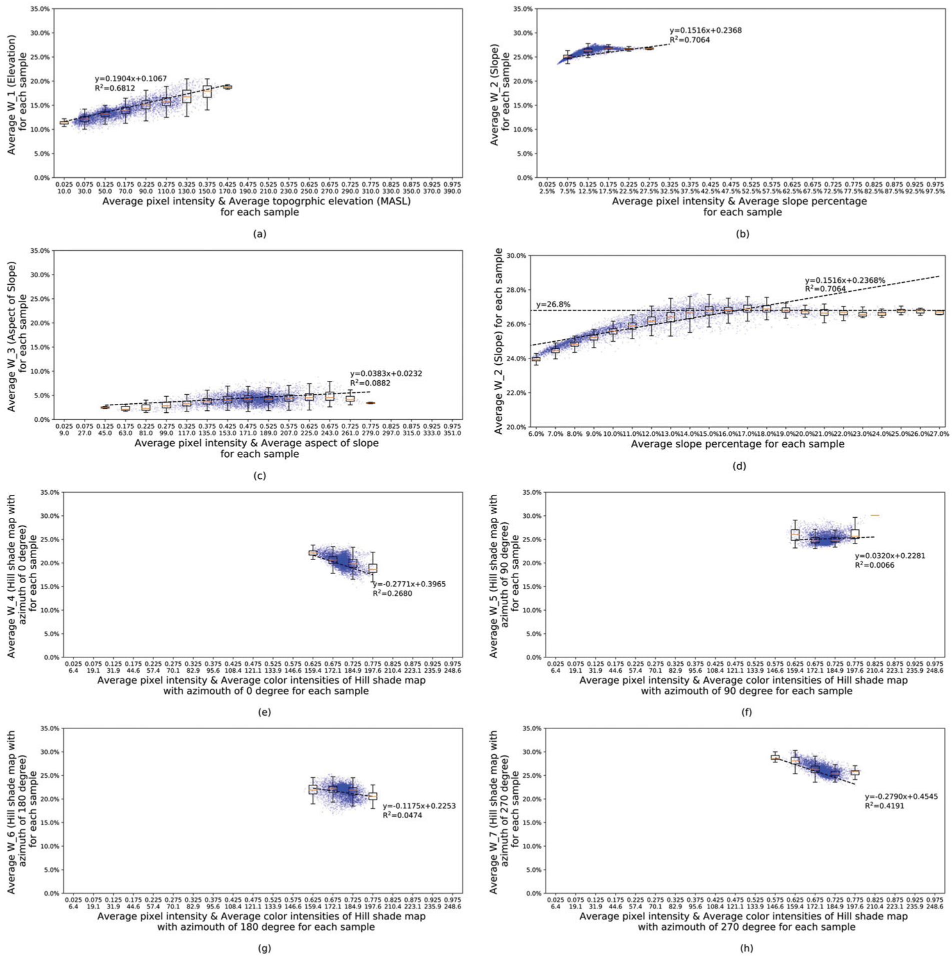

Figure 12. Correlations between channel-wise average weights for the seven types of topographic information derived from the GE module and corresponding input topographic data.

adversarial network (GAN) and can equip this model with a self-adaptive generation system for both flat and hilly urban areas.

Regarding the model structure, the proposed street network generation system extends the existing end-to-end model to a three-stage one. With the same model inputs of topographic information and contextual guidance, the extended model can use coarse road networks and their vertices as intermediate results to improve prediction accuracy. The improvements are accomplished by introducing a geo-extractor (GE) module and a geo-merging (GM) bypath. The GE module learns dynamic topographic information selection and weighting mechanisms at different spatial locations and outputs the processed topographic data to inform the system regarding the extent to which local topography should be considered in the generation tasks. The GM bypath builds on the local topographic information and junction guidance to derive vertices heatmaps that support adaptive street network generation.

The experiments for the case study areas in Italy demonstrate that the incorporation of topographic information into the generation model yields a more realistic prediction of street configurations. The model's effectiveness is further enhanced when generating streets in hilly areas. The mean absolute error and mean squared error decrease from 0.1248 and 0.2453 in the conventional Model 1 to 0.1199 and 0.2345 in the improved Model 3. Moreover, the topography-aware model (Model 3) can sharply narrow the performance differences between flat and hilly areas (from 17.02% to 3.56% in APEMC).

The model results provide insights from the perspective of computer vision to assist geographers and planners in recognizing when topographic information should be considered and which topographic data should receive more attention in road network design. Specifically, elevation and slope information should receive considerable focus in plan-making processes as their importance increases significantly in hilly urban areas. According to the GE analysis, a 9% slope percentage (5.14°) proves to be a threshold above which additional attention should be devoted in street networks design.

This research has several limitations, the consideration of which will shape the future research agenda. First, the model's generative capability is applied to areas where road networks are reasonably developed considering the surrounding context and the local topographic features. The training datasets only include two Italian cities (i.e. Florence and Perugia) where the road networks have been developed through the centuries. We acknowledge that there are also several global cities (e.g. San Francisco) where orthogonal street grids have been widely adopted even in hilly areas. Unsurprisingly, the current model will poorly predict their street networks. The varying weights of the factors determining street network design globally call for (1) the building of a larger road network training dataset that can cover more contemporary city regions, and (2) the combined consideration of both topographic and non-topographic factors in road-network generation. The enriched dataset and model inputs could help to train the GE module to learn the relationship between topography, non-topographic conditions, and the road network in various local contexts. In this way, rules of thumb that are employable globally and context-specific guidance (e.g. the slope threshold) can be derived to inform road network design.

Second, the consideration of the determinants of street configuration in this research is not comprehensive. Due to the limitations of prevalent computer vision techniques to mine the relationships across discrete features, polyline-based hydrologic features such as waterways and drainage channels are not included in the model training. Future research can employ other deep-learning techniques (e.g. graph convolutional networks) to investigate interrelationships using vector-based datasets. Moreover, many other factors might impact the hierarchical layout of roads in addition to topographic and hydrologic features, such as soil types, legal regulations (e.g. non-development area), and planning guidance (e.g. the density of buildings). Future work should introduce a combination of different determinants to clarify the relationships and their relative importance, which can better imitate plan-making processes in practice.

Notes

1. Note that the impacts of planning guidance have been explored in Fang et al (2021) and therefore is out of the scope of this research.
2. More test results can be found in Supplemental Material D.
3. Specifically, we cropped one data sample from the multi-channel base maps of Rome and Siena every 25 pix in both horizontal and vertical direction, forming a new dataset with 6,800 samples for GE analysis.

Acknowledgments

The authors are also grateful for the anonymous reviewers and editor for their constructive comments, which helped improve the quality of the paper.

Disclosure statement

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

Data and codes availability statement

The data and codes that support the findings of this study are available at the following link: https://doi.org/10.6084/m9.figshare.19606960.v1.

Funding

Zhou Fang appreciates the support from the China Scholarship Council and the Cambridge Commonwealth, European & International Trust through a CSC–Cambridge Trust scholarship (CSC No. 201808060082), Magdalene College via Ng Fourth-Year PhD Bursary, and the Alibaba Group through the Alibaba Research Intern Program. Ying Jin wishes to acknowledge funding support from the Tsinghua University Initiative Scientific Research Program via the Tsinghua–Cambridge research collaboration. Tianren Yang wishes to thank the support from the HKU-100 Scholar Fund.

Notes on contributors

Zhou Fang received his B.Eng. and M.Phil. degrees from the University of Hong Kong and the University of Cambridge, respectively. He is currently pursuing his Ph.D. degree at the Martin Centre for Architectural and Urban Studies, Department of Architecture, the University of Cambridge. His current research interest lies in deep learning and its applications in the fields of geographical information science and urban studies.

Jiaxin Qi received his B.Eng. degree from the University of Science and Technology of China in 2018. He is currently pursuing a Ph.D. degree at the School of Computer Science and Engineering, Nanyang Technological University. He is interested in the theory of deep learning and computer vision.

Lubin Fan is a Senior Algorithm Engineer at Alibaba Damo Academy with research interests including urban reconstruction, procedural modelling, image and video processing, and geometric processing. He received his B.Sc. and Ph.D. degrees in Mathematics from Wuhan University in 2009 and Zhejiang University in 2014, respectively. From 2014 to 2017, he was a postdoctoral fellow at the Visual Computing Center, King Abdullah University of Science and Technology (KAUST).

Jianqiang Huang is currently the Director of the Alibaba DAMO Academy. His research interests focus on the visual intelligence in the city brain project of the Alibaba Group. He was a recipient of the Second Prize of the National Science and Technology Progress Award in 2010. and owned the title of 'Outstanding Scientific and Technological Worker' awarded by the Chinese Institute of Electronics in 2021.

Ying Jin is a Professor of Architecture and Urbanism at the Department of Architecture at the University of Cambridge. He is the Leader of the Cities and Infrastructure Research Group and the lead convenor of the international symposium on applied urban modelling since 2011.

Tianren Yang is an Assistant Professor in the Department of Urban Planning and Design at the University of Hong Kong. He is interested in developing advanced urban analytics and modelling to decipher how city regions evolve in terms of land development, human activity, and social welfare. He earned his Ph.D. in applied urban modelling from the University of Cambridge and is a chartered urban planner in the UK and China.

References

Aliaga, D.G., Vanegas, C.A., and Benes, B., 2008. Interactive example-based urban layout synthesis. *ACM Transactions on Graphics*, 27 (5), 1–10.

Beneš, J., Wilkie, A., and Křivánek, J., 2014. Procedural modelling of urban road networks. *Computer Graphics Forum*, 33 (6), 132–142.

Boeing, G., 2019. Urban spatial order: street network orientation, configuration, and entropy. *Applied Network Science*, 4 (1), 1–19.

Chen, G., *et al.*, 2008. Interactive procedural street modeling. *ACM SIGGRAPH 2008 papers*. 1–10.

Coons, S.A., 1964. An outline of the requirements for a computer-aided design system. *Simulation*, 2 (2), R-2–304.

Fang, Z., *et al.*, 2020a. "Reading" cities with computer vision: a new multi-spatial scale urban fabric dataset and a novel convolutional neural network solution for urban fabric classification tasks. *In*: Proceedings of the 28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '20), Seattle, WA, USA 3-6 November 2020, 507–517.

Fang, Z., Jin, Y., and Yang, T., 2021. Incorporating planning intelligence into deep learning: A planning support tool for street network design. *Journal of Urban Technology*, 1–16.doi:10.1080/10630732.2021.2001713.

Fang, Z., Yang, T., and Jin, Y., 2020b. DeepStreet: A deep learning powered urban street network generation module. *arXiv preprint arXiv: 2010.04365*. doi:10.48550/arXiv.2010.04365.

Galin, E., *et al.*, 2010. Procedural generation of roads. *Computer Graphics Forum*, 29 (2), 429–438.

Hartmann, S., *et al.*, 2017. Streetgan: Towards road network synthesis with generative adversarial networks. *In:* Proceedings of the 25th International Conference in Central Europe on Computer Graphics, Visualization and Computer Visionin, 133–142.

Iizuka, S., Simo-Serra, E., and Ishikawa, H., 2017. Globally and locally consistent image completion. *ACM Transactions on Graphics*, 36 (4), 1–14.

Jo, Y., and Park, J., 2019. Sc-fegan: Face editing generative adversarial network with user's sketch and color. *In:* Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 1745–1753.

Kelvin, L.Z., and Anand, B., 2020. Procedural Generation of Roads with Conditional Generative Adversarial Networks. In: *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, 277–281.

Kempinska, K., and Murcio, R., 2019. Modelling urban networks using Variational Autoencoders. *Applied Network Science*, 4 (1), 1–11.

Law, S., *et al.*, 2020. Street-Frontage-Net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34 (4), 681–707.

Law, S., and Neira, M., 2019. An unsupervised approach to geographical knowledge discovery using street level and street network images. *In:* Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, GeoAI, 56–65.

Miyato, T., *et al.*, 2018., Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*.

Olaya, V., 2009. Basic land-surface parameters. *Developments in Soil Science*, 33, 141–169.

Parish, Y.I.H., and Müller, P., 2001. Procedural modeling of cities. In*:* Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 301–308.

Shi, W., *et al.*, 2021. RegNet: a neural network model for predicting regional desirability with VGI data. *International Journal of Geographical Information Science*, 35 (1), 175–192.

Yang, T., 2020. Understanding commuting patterns and changes: Counterfactual analysis in a planning support framework. *Environment and Planning B: Urban Analytics and City Science*, 47 (8), 1440–1455.

Yang, T., *et al.*, 2019. Aspirations and realities of polycentric development: Insights from multi-source data into the emerging urban form of Shanghai. *Environment and Planning B: Urban Analytics and City Science*, 46 (7), 1264–1280.

Yu, J., *et al.*, 2018., Generative image inpainting with contextual attention. *arXiv preprint arXiv:1801.07892*.

Yu, J., *et al.*, 2019., Free-form image inpainting with gated convolution. *arXiv preprint arXiv:1806.03589*.