InstantCITY: Synthesising morphologically accurate geospatial data for urban form analysis, transfer, and quality control

Abraham Noah Wu[a], Filip Biljecki[a,b,*]

[a]*Department of Architecture, National University of Singapore, Singapore*
[b]*Department of Real Estate, National University of Singapore, Singapore*

Abstract

Generative Adversarial Network (GAN) is widely used in many generative problems, including in spatial information sciences and urban systems. The data generated by GANs can achieve high quality to augment downstream training or to complete missing entries in a dataset. GANs can also be used to learn the relationship between two datasets and translate one into another, e.g. road network data into building footprint data. However, such approach has not been developed in the geospatial and urban data science context, its usability remains unknown, and the methods are not fully developed. We develop a new Geographical Data Translation algorithm based on GAN to generate high-resolution vector building data solely from street networks, which may be used to predict the urban morphology in absence of building data, also enabling studies in unmapped or undermapped urban geographies, among other advantages. Experiments on 16 cities around the world demonstrate that the generated datasets are largely successful in resembling ground truth morphologies. Thus, the approach may be used in lieu of traditional data for tasks that are often hampered by lack of data, e.g. urban form studies, simulation of urban morphologies in new contexts, and spatial data quality assessment. Our work proposes a novel rapid approach to generate building footprints in replacement of procedural methods and it introduces a new intrinsic method for large-scale spatial data quality control, which we test on OpenStreetMap by predicting missing buildings and suggesting the completeness of data without the usually required authoritative counterparts. The code, sample model, and dataset are available openly.

Keywords: Deep learning, machine learning, Volunteered Geographic

1

1. Introduction

The rapid growth of geospatial data in the past decades have accelerated our understanding of the world. Insights and products derived from geospatial data are solving a myriad of complex problems down the gamut from increasing fuel efficiency, monitoring logistic networks, to sustainability-driven urban designs (Lee and Kang, 2015; Lipson et al., 2022; Wagner et al., 2022; Arribas-Bel and Fleischmann, 2022; Wang et al., 2022; Li et al., 2022b). The datasets that enabled these advancements are mostly collected using a variety of methods such as surveying, remote sensing, mobile mapping, and crowdsourcing (Li et al., 2016; Huang and Wang, 2020; Jin et al., 2022; Heikinheimo et al., 2020; Luo et al., 2022; Yan and Huang, 2022).

Today, such datasets are more abundant than ever, but their quality and coverage can still vary dramatically. This phenomenon is described as Geospatial Data Asymmetry in which some geospatial data are mapped more extensively than another *correlated* data (Wu and Biljecki, 2022). This problem is more acute in open access and volunteered platforms such as OpenStreetMap (OSM). For example, more than 80% of roads has been mapped in OSM (Barrington-Leigh and Millard-Ball, 2017). However, building data, which is usually strongly associated with street networks, is estimated to be disproportionately less complete (Biljecki, 2020; Yeboah et al., 2021; Leonard et al., 2022), hampering their applications in many areas around the world.

Such asymmetry exists due to the fact that it is more complex and time-consuming to map certain features such as building footprints than tracing street networks. Automated approached used in research and industry also face similar limitations despite recent advances (Sun et al., 2020).

One novel take on the issue of data asymmetry is the introduction of Geographic Data Translation (GDT). Instead of simply collecting data from the real world, GDTs use the problem of data asymmetry to their advantage, generating a less abundant geospatial dataset by learning associations from another more available dataset. For example, Wu and Biljecki (2022) have developed GANmapper, an approach that uses Generative Adversarial Networks (GAN) to

---

*Corresponding author

Email addresses: `abrahamwu@u.nus.edu` (Abraham Noah Wu), `filip@nus.edu.sg` (Filip Biljecki)

transform one spatial dataset to another without the need of any other external data sources or on-site surveys. The model ingests the street network data in the form of XYZ tiles, a popular format used for raster maps, and generates building footprints in the same (image) format. The predicted building footprints from the deep neural network are visually similar to actual (ground truth) building footprints and achieve a high visual similarity score measured by Frechet-Inception Distance (FID), a common visual similarity metric applied to measure the performance of generative algorithms such as GANs.

While the results of GANmapper are visually compelling, we have identified two major drawbacks, which limit its applicability, and require methodological and other advancements to advance the idea and state of GDT. First, the output resolution is constrained, impeding applications that require a truthful and detailed representation of the built environment. In situations where building footprints are small and closely packed, it is common to observe irregular shapes and multiple buildings blended together, reducing the accuracy and detail, and limiting the use of data. Methodological advancements are necessary to solve these shortcomings. Second, its applications are limited and not developed, e.g. the model only creates data as raster tiles and struggles with converting them into geo-referenced vector data. While the tile/raster format is expedient for online viewing, quantitative (e.g. cartographic) metrics such as site cover, area difference, and perimeter difference, which are used for a plethora of spatial analyses, cannot be directly calculated from this raster format. This shortcoming severely hinders the statistical understanding and actual application of such results, and overcoming it requires further advancement and expansions of the methodology.

In this paper, we introduce InstantCITY, a new GDT method that can generate up substantially higher (4×) resolution results and accurate vector representations that are directly applicable in multiple use cases, which we also investigate in the work for the first time.

In the experiments, the model exhibits high accuracy (e.g. a mean site cover difference of -2.5% compared to the ground truth), providing reliable results to real-world uses. Further, it demonstrates high flexibility and adaptability when applied to different urban typologies around the world. Namely, the models trained with datasets from one city behave well when transferred to cities having similar morphologies. In addition, between pairs of cities with distinct morphologies, the models double as style transfer algorithms, transferring the style in which it was trained into the target city.

We are also able to demonstrate that the model, which we release openly, is able to generate outputs that are sufficiently accurate that they can even be used

3

as reference data for spatial data quality control. For example, we demonstrate that the method excels at classifying the completeness of built up areas (e.g. identifying those that are undermapped), achieving a mean accuracy score 92.4% by comparing the synthetic data with OpenStreetMap data of two well-mapped cities, in which we have simulated various levels of degraded completeness to test the reliability of the approach.

2. Background

2.1. Introduction to Generative Adversarial Networks

Generative Adversarial Networks (GANs) is a type of generative algorithm that pits two neural networks in contest with each other in a zero-sum game to generate synthetic data that is close to the data distribution of the original dataset (Goodfellow et al., 2014).

Before GANs, generative models were unsupervised where the model would seek to learn the regularities in the input data in such a way that it can be used to generate new examples that closely represent the input data (Han et al., 2018; Chen et al., 2018; Maaløe et al., 2016). This type of unsupervised generation is often time-consuming and yielded unsatisfactory results. GANs, on the other hand, added a supervised task to the generative process. During training, the generator component in the GAN must compete with the adversarial discriminator to produce samples that could pass its scrutiny as real samples while the discriminator network must actively learn to classify the synthetic samples from the real samples. As the training process converges, the synthetic samples produced by the generator would become so close to the real dataset that the discriminator would only have a 50% accuracy, signalling that the generated dataset is now indistinguishable from the real by the discriminator (Goodfellow et al., 2014, 2016).

This type of generative algorithm has achieved many state-of-the-art results in generative problems, especially in image generation. By introducing Convolutional Neural Networks (CNN) to the architecture (Radford et al., 2015), GAN can generate high resolution synthetic photos of human faces, street view images and satellite images that could pass as real ones to human perception (Brock et al., 2018; Zakharov et al., 2019; Karras et al., 2021; Zhao et al., 2021; Toker et al., 2021; Biljecki and Ito, 2021). Equipping GANs with sequential architectures like Long Short Term Memory layers (LSTM), it can generate time-series data such as stock market trends, electrocardiograms and electricity consumption that replicates the original data without disclosing privacy information (Zhu et al., 2019; Yoon et al., 2019) and music sequences from lyrics and single latent

vector (Yu et al., 2021; Engel et al., 2019). Besides generating data, GANs are also useful in data upsampling, data privacy protection, and data extrapolation as they excel at learning and reproducing the ground truth. This trait has made GANs applicable in various domains such medicine (Beaulieu-Jones et al., 2019; Litjens et al., 2019; Bowles et al., 2018) and the built environment (Quintana et al., 2020; Yan et al., 2020a; Rachele et al., 2021).

GANs can also take in external conditions to allow certain control of the generated data. In this way, users would be able to influence the generative process to obtain the desired results. For example, some GAN architectures allow granular semantic control of the generated dataset (Isola et al., 2017; Park et al., 2019). During training, a mask-image pair is used as input in which certain image semantics are correlated to a specific semantic class. Therefore, the model would be able to learn the weights to translate simple color-coded masks to realistic-looking images, or translate photos into different stylistic expressions (Zhu et al., 2017). Using a similar concept, GANs can be used in content-aware image inpainting to fill up missing areas in an image (Li et al., 2017; Yeh et al., 2017; Pathak et al., 2016).

2.2. Applications of Generative Adversarial Networks in GIS

With the power of GANs verified by the scientific community, GIScience researchers have also begun investigating the applications of GANs in GIS and related domains (Wu et al., 2022; Quan, 2022). GANs have been used as a cartography tool to generate synthetic satellite images of landscapes and cities (Abady et al., 2020; Zhao et al., 2021), transfer between cartographic styles (Christophe et al., 2022), translate satellite images into cartographic representations (Isola et al., 2017; Li et al., 2020b), and generate cartographic representations from geospatial vector data (Kang et al., 2019). They are also used to create semantic-responsive land-cover maps with user-drawn colour masks (Park et al., 2019; Baier et al., 2021). Andrade and Fernandes (2020) also explored translating historical maps to satellite images to raise the awareness of shifting landscapes.

However, most applications of GANs in spatial information sciences stop at the novelty of generating raster maps without tackling more pressing issues in the field such data accessibility, data accuracy, and data asymmetry. There are currently only a few ventures in using GANs for data augmentation. Zhang et al. (2021b) revealed that GANs can be used to restore deformed satellite images caused by jitters in geopositioning. GANs have also set the state-of-the-art performance in satellite image sharpening and superresolution (Ma et al., 2020; Jiang et al., 2019). In addition, GANs can also be used to remove cloud contamination in remote sensing images (Li et al., 2020c), while Zhang et al. (2021a)

5

have developed a method that could automatically remove unwanted objects in street view imagery.

More recently, GANmapper (Wu and Biljecki, 2022) introduced the idea of Geographic Data Translation (GDT) and demonstrated that GANs can ingest street network data and generate (inpaint) building footprints in raster format that visually look realistic. This technique is especially useful in addressing the problem of data asymmetry (i.e. the difference in data quality between spatially correlated datasets), which has always been a pain point in Volunteered Geographic Information (VGI), e.g. OpenStreetMap (OSM) (Li et al., 2020a).

Besides road-building data pair, other asymmetries in OSM data also exist. For example, there remain places with undermapped road networks and primary roads can be used to infer secondary and tertiary roads. Furthermore, areas with building height information can be learned by the model to generate approximations of building height data in another with similar morphology (Milojevic-Dupont et al., 2020).

By translating a more readily available geospatial dataset into a less available dataset, the method can greatly scale up the mapping of complex datasets. However, as indicated in the introduction, there are still major limitations to this method, namely, the low output resolution and the limitation of a raster output. Therefore, additional effort needs to be done before GDT models can find practical use in GIS, urban form studies, cartography, and related domains, inviting a new array of applications beyond merely visual output.

2.3. Procedural Urban Modelling

As hinted at in the introduction, the method introduced in this paper is able generate building data that is both visually realistic and sufficiently accurate. The results closely resemble the urban form of the real-world situation in the absence of conventional building data. Therefore, it can serve as an more scalable alternative technique to procedural modelling, which warrants a brief overview.

Procedural modelling is a type of generative modelling using a set of rules coded by designers which can replace the cumbersome process of manual modelling. It is widely used in creating architecture and cityscapes in films and games (Smelik et al., 2014; Tobiáš and Cajthaml, 2020). In the built industry, architects and planners can also use procedural models for rapid prototyping of architectural facades or city layouts, evaluating different design options without having to manually reconfigure the digital models (Fink and Koenig, 2019; Groenewegen et al., 2009; Kim et al., 2018; Birch et al., 2001). However, like all other heuristic algorithms, the procedural models need to coded by experienced designers and has certain degree of repetition in the generated cityscape

as the randomness of the results are based on the complexity of the programme. Furthermore, when procedural modelling is used to simulate an existing setting, the results are not necessarily accurate and they may not reflect the real-world reliably. Other the other hand, GDT models based on GANs can learn the intricacies of the spatial relationship from a large sample size and do not require designers to program the rules behind a specific morphology and is more stable and smarter in adapting to cities with different road typologies.

We show in this paper that models trained with datasets from one city behave well when transferred to cities having similar morphologies. Between cities with distinct morphologies, the models act as style transfer algorithms, transferring the style in which it was trained into the target city.

2.4. *Spatial Quality Assessment*

Spatial data quality is pivotal across numerous fields, and in GIScience is a research line on its own with a mountain of published research papers (Li et al., 2022a; Senaratne et al., 2016; Hou and Biljecki, 2022). In this paper, we take advantage of the generated data to study the quality of another dataset, introducing a new method for spatial data quality assessment. We focus on one of the most important quality indicators — the completeness of the dataset — the proportion of the features that are mapped with respect to the real world, i.e. understanding whether an area has been mapped entirely.

There are various approaches for spatial data quality assessment developed so far. As the quality of urban data is becoming increasingly important (Basiri et al., 2019; Songchon et al., 2021; Grinberger et al., 2021), various methods to assess the completeness of features have been developed, with many of them focused on buildings (Senaratne et al., 2016). There are intrinsic methods, i.e. predicting the completeness of features based on the history of contributors or the arrangement of existing features (Zhou, 2017; Jacobs and Mitchell, 2020; Majic et al., 2021; Sundaram et al., 2021), and those that are extrinsic, requiring checking against another, usually authoritative, dataset representing the same features or proxies (Brovelli et al., 2016; Balducci, 2019; Li et al., 2020a).

In the experiments, we investigate whether our method can also be used as a key component in spatial data quality assessment. Using the synthetic dataset trained on high quality data as reference, we research whether our method can be used as a new approach in the spatial data quality assessment research to sense completeness of buildings without having to rely on often unavailable authoritative (external) data.

3. Methodology

3.1. Overview

We propose a new model architecture to enable training and inference of urban data at high-resolution per tile using street networks input as images. Thanks to the high resolution output, the raster result can be vectorised into individual building polygons to enable downstream geospatial applications and validation of the method using geospatial metrics on top of the ones used in computer vision (Figure 1).

InstantCITY Methodology

Slice → Generate → Stich → Vectorize

Raster Map — Input Tiles — Output Tiles — Completed Raster Map — Georeferenced Vector Data

Completed Raster Map
1024*1024px
FID: 76.2

Georeferenced Vector Data
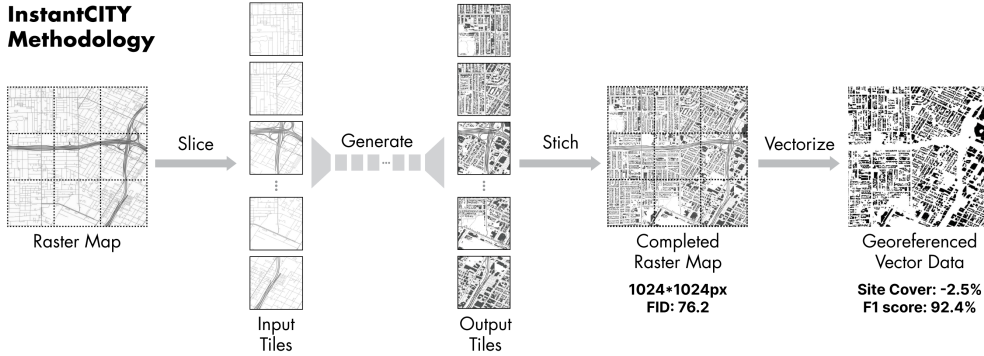**Site Cover: -2.5%**
F1 score: 92.4%

Figure 1: InstantCITY visual summary.

To investigate the performance and flexibility of our model, we used OSM data of 8 cities across the world, representing a range of urban morphologies. These cities are major cities in the world that generally have a high degree of building data completeness to provide adequate training data and ground truths for evaluation.

In the experiments, we will examine the performance of the model at different zoom levels and in different cities using both visual and GIS metrics and explore two potential applications of the methodology in urban design and data quality control. Furthermore, we also investigate how the method generalises from one city to another, enabling the development of the method in a well-mapped area and applying it in an urban environment with data of deficient quality.

3.2. Model Architecture

The InstantCITY model architecture is a type of Image-to-Image Conditional GAN (Mirza and Osindero, 2014; Isola et al., 2017) that translates input image data such as street networks to a target image populated with generated building footprints.
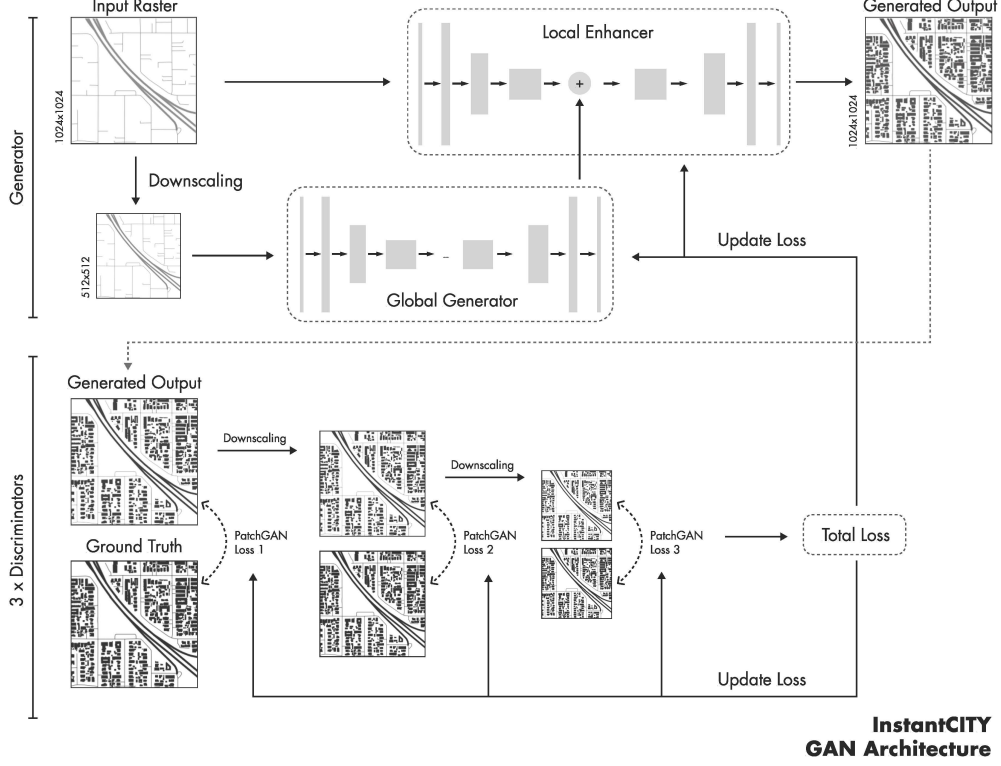
8

Figure 2: InstantCITY GAN Architecture.

The ability of InstantCITY to generate high-res images at 1024x1024 pixels is achieved using a Coarse-to-fine generator (Wang et al., 2018) that generates target images at different scales (Figure 2). To achieve this performance, two residual networks G1 and G2 are constructed. The model is trained on the residual network (He et al., 2016) G1 on lower resolutions first and the G2 is appended to G1 to be trained jointly on high resolution images. The overall structure of the generator is an encoder-decoder pair (Hinton and Salakhutdinov, 2006) where the first half of the model encodes inputs into latent vectors and the latter part decodes the vectors into visual representations.

In each forward pass, the generator will try to generate outputs that could 'fool' the discriminator into classifying the generated image as 'real', while the discriminator will learn to classify the generated images as 'fake' and ground-truth targets as 'real'.

There are three discriminators at different resolutions to determine the quality of the output image at different resolutions. In each discriminator, Convolutional

9

Kernels are applied to extract and compare the features from both the generated images and the ground truths. The final classification is the average value across 3 discriminators that indicate whether the generated image is real or fake.

At the end of each forward pass, the loss for both the generator and discriminator is calculated and their weights updated. As each epoch passes, both the generator and discriminator get better in their roles until the generator's output would pass the scrutiny of the discriminator as real images 50% of the time, hinting that both the generator and the discriminator have converged to the best possible performance with no one winning against the other.

3.3. Pre- and post-processing

As with other deep learning models, the resolution of the input and output is limited by the memory of the hardware, and the resources needed scale proportionally to the size of the input images. Thus, it would be too resource-intensive to fit a large area (e.g. an entire city or a neighbourhood) within a single high-definition image. Instead, we use a pre-processing pipeline to convert the targeted areas into 1024x1024 raster tiles into a WMTS (XYZ tiles) directory similar to previous works in Geospatial Deep Learning (Ng and Hofmann, 2018; Wu and Biljecki, 2021). The output tiles are generated according to the same directory structure as the input tiles and can be restored into a larger whole. The raster outputs are then stitched to into a large area and the raster tiles are converted into vector data for geospatial analyses against the ground truth data.

According to Wu and Biljecki (2022) and Chen et al. (2021), 3-channel RGB images enhance the performance of residual networks in both generative and classification problems. In our case, the colour and the width of the lines are also used to identify the type of streets. For example, highways are red with a thicker line width and secondary roads are blue with a thinner line width. The output images also contain the street networks to stabilise the training and help the model to converge quicker. The different colour coding of street networks and building footprint also makes extraction of footprints and subsequent vectorisation easier.

3.4. Metrics for evaluation

Frechet Inception Distance (FID) (Heusel et al., 2017) is the standard benchmark for measuring the performance of GANs. FID represents the distance between the feature vectors of real and generated images that are computed using an Inception-v3 (Szegedy et al., 2016) image classification model.

A lower FID score indicates that the two groups of images are more similar in terms of the extracted feature vectors. A perfect score of 0.0 indicates that the two groups are identical. Visually, a lower FID score on the generated images

tends to correlate well with realistic-looking images, indicating more morphological correct representations in terms of the general shape, size, and density of the generated building footprints. Mathematically, FID can be expressed with the following formula:

$$FID = \left\| \mu_r - \mu_g \right\|^2 + Tr(\sum r + \sum g - 2\sqrt{\sum r \sum g})$$

where $X_r \sim N(\mu_r, \sum r)$ and $X_g \sim N(\mu_g, \sum g)$ are the 2048-dimensional activations of the Inception-v3 pool3 layer for real and generated samples respectively.

In addition to FID, the Mean Intersection over Union (mIoU) between the inputs and generated images is also calculated in some experiments to measure the degree of overlap of the generated images to the ground truth. Mathematically, mIoU measures the number of pixels common between the input and generated images divided by the total number of pixels present across both datasets.

A high mIoU score indicates that the generated images overlaps with the input more and thus replicates the input data better. However, due to the generative nature of GANs, the generated results should look realistic in terms of the overall building morphology but not overlap exactly with the input image. If the generated image has a high mIoU score (close to 1), we can conclude that the generator is overfitted to the training data, which impedes the models' ability to generalise its learned patterns to new inputs.

3.5. Experiment Setup

Four experiments are set up to investigate the performance and potential application of our model. Experiments 1 and 2 look into the performance of the model in 8 cities at two scales (1000m per tile and 500m per tile). Both visual and statistical metrics are used to assess the differences between the generated data and the ground truths to understand the capabilities and the limitations of the method. Experiment 3 investigates the ability of the model to apply its learned parameters in a new city, and essentially its capability to serve as a means for style-transfer. We selected 4 models and generated the building data of 8 target cities (elaborated in the next section). Some of the cities have similar morphologies to the one where the models are trained, whereas some target cities have have drastically different morphologies. This comprehensive investigation helps to shed insight into how the trained models perform under different conditions, and this experiment also doubles in understanding whether the method can be used in urban planning research to transfer the morphology from one city to another. Experiment 4 investigates how the model could be used as a tool for OSM quality control, essentially introducing a novel approach in assessing the quality

of spatial data, especially of Volunteered Geographic Information (VGI) (Yan et al., 2020b), which are burdened with lack of authoritative data at the global scale. The existing OSM datasets of multiple cities (attested to be complete) are disturbed with random errors, reducing their completeness to create an artificially incomplete dataset. This degraded dataset is compared with the dataset generated by our method and its performance in classifying under-mapped areas in a large region can be tested.

3.6. Implementation

The model is implemented with PyTorch (Paszke et al., 2019), and QGIS and Mapbox API is used in data processing. We release InstantCITY as open-source software. Further, in the repository, we also include the pre-processing and post-processing pipelines, which may aid other researchers to extend the work. These are available at `https://github.com/ualsg/InstantCITY`.

4. Results and evaluation

4.1. Experiment 1 — Investigating model performance

To evaluate the accuracy and versatility of the model, we picked 8 cities with different morphologies and trained different models for each of the cities. The cities are Beirut, Frankfurt, Jakarta, London, New York City, Rotterdam, Seattle and Singapore. The dataset of each city is obtained from OSM and the raster tiles are generated at two zoom levels representing 500 and 1000 meters per tile, which approximates to 0.5 meters per pixel and 1 meter per pixel, respectively.

Figure 3 illustrates the generated images from a subset of cities with the pair of input images and real images. The colour coding of the street network helps the model to learn its network hierarchy, and the grey patches represent water body which helps to differentiate the boundary between land and water. The images chosen in this figure represent different urban morphologies: Seattle has large, rectangular blocks with standalone houses; Frankfurt has a courtyard typology with buildings facing the street and back-of-houses within each block; Jakarta is populated with small, ad-hoc structures that could become extremely dense; and Rotterdam has a similar courtyard typology as Frankfurt, albeit with much more regular blocks. As shown in the figure, the model is able to generate realistic representations of the real image in most scenarios with the exception of Frankfurt. This might be due to the fact that the real image has many complex building polygons within the courtyards that are not guided by any street network, therefore the model is unable to infer some of the shapes within the

Figure 3: Results from 4 different cities with different morphologies at 1000m/tile zoom level.

courtyards. However, in all cases, the model is able to generate images that represents the real building footprints accurately in terms of granularity and urban texture.

Figure 4 shows additional results from the other four cities included in the experiments. Models can be trained at two zoom levels for each city. At 1000m zoom level, each tile covers the size of a neighbourhood and could capture a variety of typologies in one tile. This result suggests that the model is robust enough to infer the shifts in building typologies in different parts of a city. At the 500m level, each tile covers a few blocks and is generally homogeneous in terms of building typologies. At this level, the model is able to generate images with sharper corners that represent the real images in a higher degree.

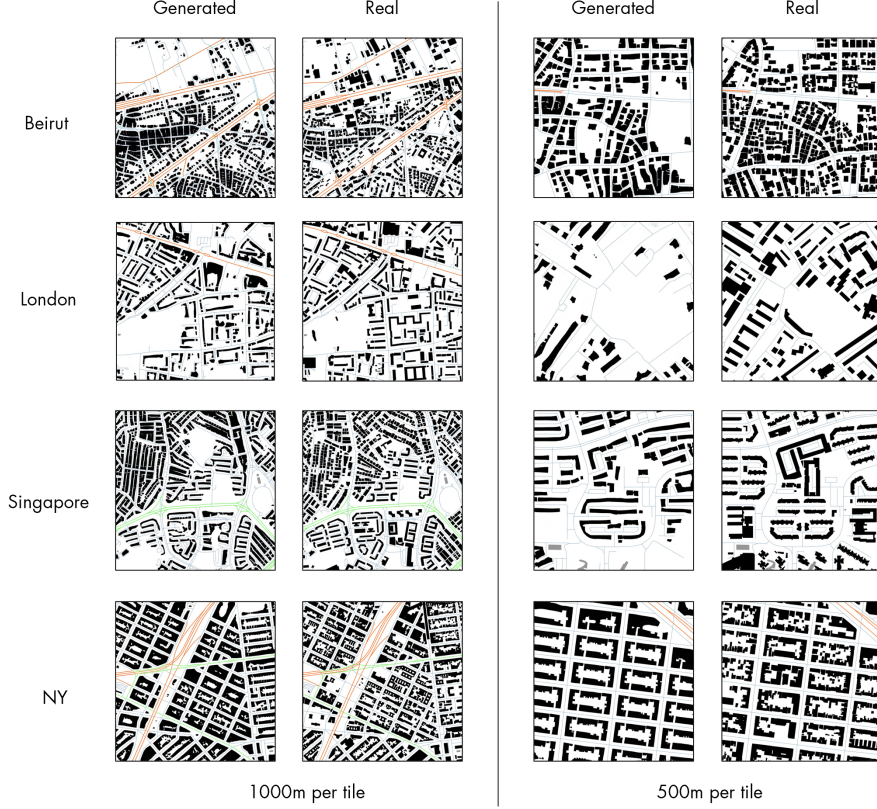The visual closeness between the ground truths and the generated imaged

Figure 4: Additional results at two different tile zoom levels.

images can be measured by FID scores as explained in Section 3.4. A lower FID score indicates that the generated images are closer to the real images and thus the model's performance is better and vice versa. For each model trained, the FID scores are calculated against the ground truth in two scales. The results are displayed in Table 1.

From the table, we observe that the performance of the model varies from city to city. Different urban forms would have different impact on the final output even though all models are trained with the same settings. Comparing the FID scores with the sample images from Figure 3 and 4, we can see that the scores are effective at representing the quality of the generated images. Singapore performs well compared to other cities in FID scores (65.7 at 1000m and 55.8 at 500m), and we can see from Figure 4 that the generated image is able to capture the shape, size and the overall urban form of the ground truth closely at both scales. On the contrary, while London scores well at the 1000m scale at 65.6, it slipped

Table 1: FID scores by city.

Scale	City	Training size	FID
1000m	Beirut	101	77.8
	Frankfurt	131	97.1
	Jakarta	171	63.4
	London	180	65.6
	NYC	238	84.1
	Rotterdam	165	90.0
	Seattle	91	66.2
	Singapore	233	65.7
Average			**76.2**
500m	Beirut	364	63.8
	Frankfurt	436	142.8
	Jakarta	608	61.9
	London	718	118.6
	NYC	830	57.8
	Rotterdam	609	87.7
	Seattle	289	66.3
	Singapore	781	55.8
Average			**81.8**

at the 500m scale at only 118.6. We can observe the difference hinted by the FID score in the quality of generated image in Figure 4 as well. At 1000m per tile, the generated image for London represents the ground truth very closely, while the image at 500 per tile clearly misses out on many buildings and buildings that are generated are different from the ground truths in terms of shape and size.

While the average FID scores of both scales are similar (76.2 at 1000m and 81.8 at 500m), the 500m model generated two outliers that do not produce satisfactory results. Both London and Frankfurt at 500m have much higher FID scores compared to the rest of the cities and many generated tiles do not resemble the ground truth well. This is probably due to the fact that both cities have a lot of buildings designed with an European courtyard typology which are more likely to be clipped into multiple buildings at the scale of 500m. Since the model only learns the relationship of pixels within each tile, the greater contextual information of the larger urban block is lost when zoomed in at the 500m scale. When the contextual information is preserved at the 1000m scale, the scores of both cities improved significantly with London having one of the best FID scores.

In conclusion, from both the figures and the FID score, we have shown that

the proposed model is able to generate high definition synthetic representations of building footprints that represent the ground truth very closely with the exception of 2 cities at 500m. We have also shown that FID score is a useful metric in comparing the performance of the model in different situations without manual inspection of the images.

4.2. *Experiment 2 — Applying geospatial metrics on generated data*

While most of the models trained in Experiment 1 returned visually compelling results, the exact shape and size of each generated footprint have not yet been measured. Looking back at Figure 4, the model does not always generate sharped cornered polygons that represent buildings, but a high probability region that contain a building.

To explore the performance of the model at the building level, the raster images can be vectorised into geospatial polygons to support GIS operations and computation of quantitative urban morphology metrics, and may invite several new applications of GAN-generated spatial data, some of which we investigate in this section. This will allow us to calculate metrics at the polygon level which will provide deeper insight into the performance of the model. To remove score biases caused by difference in city areas, we randomly cropped a 8x8km region from each city to calculate the statistics. Additionally, the selected region is also divided into tiles at the same scale of the model used so that vector statistics can be aggregated at the tile level as well.

To understand the potential of this method, we have used urban form metrics, e.g. site coverage ratio and average building footprint size, that are typically used in urban form studies across multiple domains (Ahn and Sohn, 2019; Zhang et al., 2019; Heris et al., 2020; Li et al., 2020d; Biljecki and Chow, 2022). Table 2 compares the differences in metrics of the generated dataset with the ground truth dataset. In most metrics, the median value is taken rather than mean value since the effect of outliers might be higher on the relatively small area of testing. Δ **Bldg. Area** (Equation 1) and Δ **Bldg. PM** (Equation 2) measures the median percentage area and perimeter difference between the generated and the ground truth of each individual polygon respectively while Δ **Site Cover** (Equation 5) measures the median percentage difference in the built area of between the sum of polygons of each tile. Thus, a generated dataset would be considered better in quality if it scores close to 0% in Δ **Bldg. Area**, Δ **Bldg. PM**, and Δ **Site Cover**, this indicates that the overall size of the prediction is similar to the ground truth. **% GN Count** (Equation 3) measures the proportion of the number of polygons in the generated set against the ground truth in percentage and **mIoU** (Equation 6) measures the intersection over union score averaged at

16

the tile level. Thus, a generated dataset that scores close to 100% or 1 in **% GN Count** and **mIoU** respectively is also considered a good prediction as the number of predicted buildings accurately portrays the number of buildings in the ground truth and a high proportion of predicted area intersects with the ground truth, indicating higher accuracy in terms of location and shape. Moreover, **Exp 1 FID** is the FID score computed in the previous section serving as a reference to other vector metrics and a lower FID score indicates a higher visual similarity of the generated dataset compared to the ground truth.

$$\Delta \text{ Bldg. Area (\%)} = 100 \times (\text{Polygon Area}_{\text{GN}}/\text{Polygon Area}_{\text{GT}} - 1) \qquad (1)$$

$$\Delta \text{ Bldg. PM} = 100 \times (\text{Polygon Perimeter}_{\text{GN}}/\text{Polygon Perimeter}_{\text{GT}} - 1) \qquad (2)$$

$$\text{\% GN Count} = 100 \times (\text{Polygons per Tile}_{\text{GN}}/\text{Polygons per Tile}_{\text{GT}}) \qquad (3)$$

$$\text{Site Cover} = 100 \times (\text{Tile Building Area}/\text{Tile Area}) \qquad (4)$$

$$\Delta \text{ Site Cover} = \text{Site Cover}_{\text{GN}} - \text{Site Cover}_{\text{GT}} \qquad (5)$$

$$\text{mIoU} = \text{Area of Intersection}_{\text{Tile}}/\text{Area of Union}_{\text{Tile}} \qquad (6)$$

Looking at the average values of the metrics for both zoom levels in Table 2, we see that models at 1000m per tile performed better in all metrics. Figure 5 shows some examples of the stitched results using data from the two scales. The visual results resonate with the tabular results as the results at the 1000m/per tile level are denser and have better morphological contiguity. This difference in performance is especially poignant in London. The 500m model failed completely for most tiles as it is unable to infer the building shapes from the input tiles while the model at 1000m is stable in the prediction and offers good quality results comparable to the performance of other cities at the 1000m scale. As briefly mentioned in the previous section, it is possible more contextual information is available at the 1000m in each tile, allowing the model to train and infer with better stability.

Looking at the differences in building area and perimeter from the table, the generated polygons tend to have larger area that the ground truth with the only exception of London having negative values. In other cities, the median polygon sizes are larger, sometimes going to 40% larger than the ground truth in the case of New York City and Rotterdam at 1000m while the number of individual polygons in a tile indicated by %GN Count is generally smaller than the ground truth. This indicates that a proportion of generated polygons are joined together as a single polygon while those should have been individual polygons. This is due to the fact that the model might struggle with edge separation of smaller buildings

Table 2: Accuracy of metrics characterising the urban form by city.

Scale	City	Δ Bldg. Area (%)	Δ Bldg. PM (%)	% GN Count	Δ Site Cover (%)	mIoU↑	Exp 1 FID ↓
1000m	Beirut	14.6	4.3	59.6	-6.6	0.27	77.8
	Frankfurt	8.6	-1.0	83.0	-2.6	0.39	97.1
	Jakarta	22.8	9.1	64.6	-5.7	0.38	63.4
	London	-20.7	-15.8	101.5	-1.8	0.45	65.6
	NYC	40.6	15.7	77.2	0.0	0.53	84.1
	Rotterdam	40.5	27.6	66.7	-1.4	0.47	90.0
	Singapore	30.8	18.0	67.7	-1.6	0.43	66.2
	Seattle	17.7	11.3	74.2	-0.6	0.40	65.7
Average		**19.4**	**8.7**	**74.3**	**-2.5**	**0.41**	**76.2**
500m	Beirut	12.2	4.2	40.1	-17.3	0.14	63.8
	Frankfurt	58.7	15.6	46.4	-7.7	0.22	142.8
	Jakarta	80.1	40.1	37.3	-10.1	0.28	61.9
	London	-79.2	-58.7	34.4	-38.9	0.04	118.6
	NYC	32.3	6.1	71.4	-0.1	0.55	57.8
	Rotterdam	43.3	22.4	53.7	-2.5	0.31	87.7
	Singapore	50.0	13.9	48.3	-6.0	0.30	66.3
	Seattle	44.4	17.5	66.5	-0.3	0.34	55.8
Average		**30.2**	**7.6**	**49.7**	**-10.4**	**0.27**	**81.8**

and the vectorization process has converted those footprints into a single, larger polygon. At the same time, in areas where the mapped polygons do not have a clear relationship with the street network, the model will not be able to generate useful information and thus Δ **Site Cover** is generally negative, indicating that the total building area in the generated dataset is lesser than the ground truth. Figure 6 gives another insight in the performance of the model. The errors are categorised by different classes of density and by cities. The distribution of errors affirms that the performance is driven by both the density and type of the urban form, e.g. the areas in which the density of buildings is low exhibits a high dispersion of errors, but a low median error.

The Mean Intersection over Union (mIoU) reveals the degree of overlapping area between the generated and the ground truth. There is no clear relationship of mIoU with the rest of the metrics. For example, while Frankfurt (mIoU 0.39) at the 1000m scale scored well in other vector metrics, it's mIoU is lower than Rotterdam (mIoU 0.47) which does not perform well in other vector metrics. The mIoU score also does not always correlate strongly with FID scores. For example, while New York City at 500m had the highest mIoU score of 0.55 and
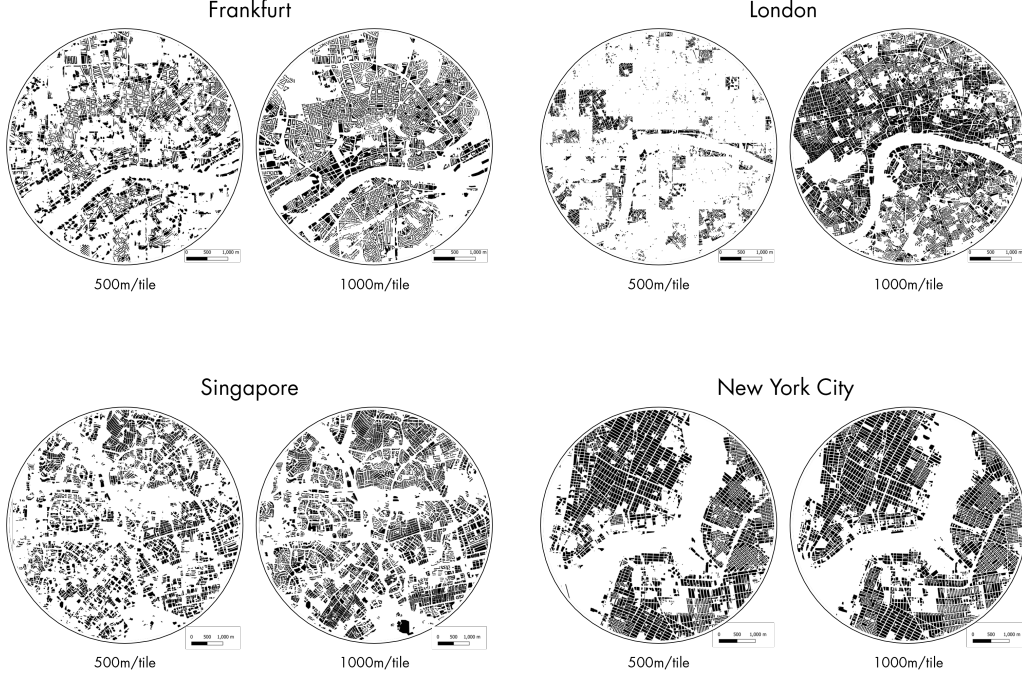
Figure 5: Selected results pairs comparing the performance of models trained at 500m/tile and 1000m/tile. The results generated by the 1000m/tile models are superior in terms of density, texture and contiguity.

came in at the second best FID score of 57.8, Seattle at 500m only scored 0.34 in mIoU score despite having the best FID score at 55.8.

This inconsistency of mIoU scores for GANs is widely discussed in related papers. Since GANs are not supervised, the generator in the model does not actually see the actually ground truth images during training. Instead, it only receives the result of the discriminator (whether the generated image is considered True according to the discriminator) during back-propagation. Thus, even though the generated data might look like the ground truth in terms of building size, density, and distribution, the specific location of the individual polygons in the generated dataset is not the objective of the learning process, resulting in inconsistent mIoU scores.

With the understanding of the tabular metrics and sample images, we can see that our model, especially trained at the 1000m scale, can generate building footprints that represent the ground truth in terms of size, density and distribution to a reasonable degree of realism. The average error in Site Cover Percentage is only -2.5% with NYC at both scales achieving 0% and -0.1% error and the
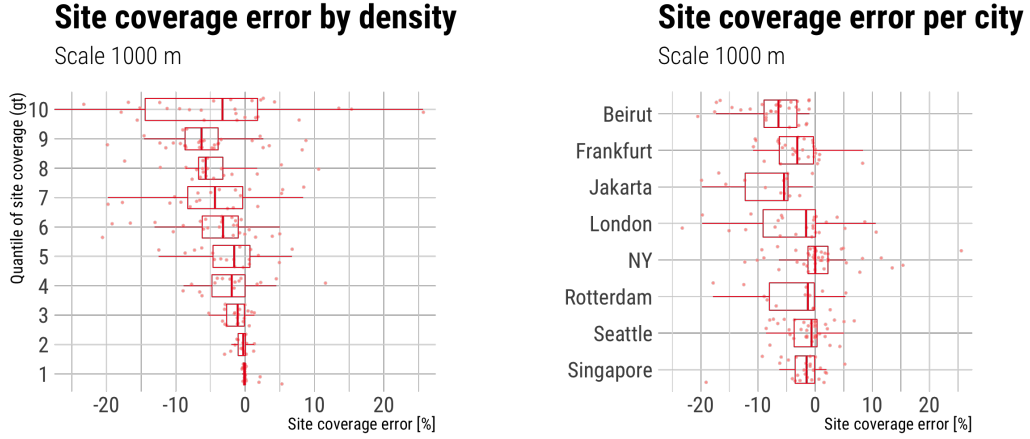
19

Figure 6: Error distribution in predicting one of the prominent metrics in urban morphology (proportion of an area covered by buildings), broken down by the density of the urban form (left), and different mix of urban forms i.e. cities (right).

median building area can be as low as 8.6%. However, it is important to note that the vectorised metrics do not necessarily correlate to visual likeliness as it is an averaged score across the test region. Both FID scores and vector metrics need to be considered when evaluating the models.

4.3. Experiment 3 — Adaptability of the model in new areas

In Experiment 2, we have shown that our model is able to generate approximations of building footprints that is visually realistic and have acceptable statistical error in terms of vectorized dataset. In this experiment, we explore the performance of transferring the morphology learned in one city to another city, which is the true test of our approach as it may enable developing the model in a well-mapped urban area and applying it in an unmapped counterpart. We hypothesise that models trained in one city could also be used in other cities as long as the target city's morphology is similar to the city in which the models are trained. For example, the model trained with the data extracted from Seattle could potentially be transferred to Chicago as both cities have orthogonal streets and similar block typologies.

In this experiment, we will test the above hypothesis and also experiment with cities having disparate morphologies to evaluate the extent of the efficacy of this approach. Table 3 outlines the results of cities using the proposed metrics in the previous section and Figure 7 shows a few examples of the generated cities.

From the table, we see that the performance of the key urban form metrics could vary vastly between cities. Correlating the tabular data with the selected

Table 3: Metrics on transfer learning for similar and disparate cities.

Model	Applied City	Δ Bldg. Area (%)	Δ Bldg. PM (%)	% GN Count	Δ Site Cover (%)	Exp 1 FID ↓
NYC	Detroit	105.9	62.9	64.1	5.3	55.9
	Jersey	59.5	27.0	56.7	-1.7	109.0
Seattle	Chicago	97.4	42.6	49.7	-2.8	113.4
	San Francisco	2.4	-8.6	93.4	-11.9	110.5
London	Manchester	-16.5	-0.6	122.7	1.0	153.6
	Paris	-76.8	-50.2	179.4	-38.6	112.4
Jakarta	Manila	-20.4	-17.3	80.5	-19.1	87.1
	Surabaya	155.7	54.9	37.4	-6.4	110.5

visual results in Figure 7, we see that the some pairs have realistically looking urban patterns and a good score.

For city pairs that have similar morphologies, the model is generally able to reproduce building patterns that represent the density and texture of the ground truth realistically, indicating that it is possible to develop a model for a series of morphologically interchangeable cities, from which those that in practice suffer from poor data availability may benefit by having InstantCITY generate truthful data that is usable for a series of spatial analyses. In the example of Seattle to Chicago, the generated result of San Francisco shows the same variation in building density from the city center to the suburbs. For city pairs that have drastically different morphologies such as in the case of London and Paris, it is expected that the results will not replicate the morphology of the ground truth, rather, the model would apply it's learned parameters onto the target city. In this example, the courtyard typology of London is transferred onto the Parisian grid and the prediction remains stable.

The behaviour of the models uncovered in this experiment could lead to applications in the urban practice as well. The trained models could be useful to help designers and planners in cross-referencing one urban morphology with another. They no longer need to 'hardcode' procedural functions to generate urban patterns for rapid prototyping. Rather, they could simply apply the models trained on a particular morphology to a region to quickly evaluate design options.

4.4. Experiment 4 — Applying model for OSM quality control

In the previous experiments, we have evaluated the performance of the model at both the polygon level and the tile level. Due to the nature of GAN, although
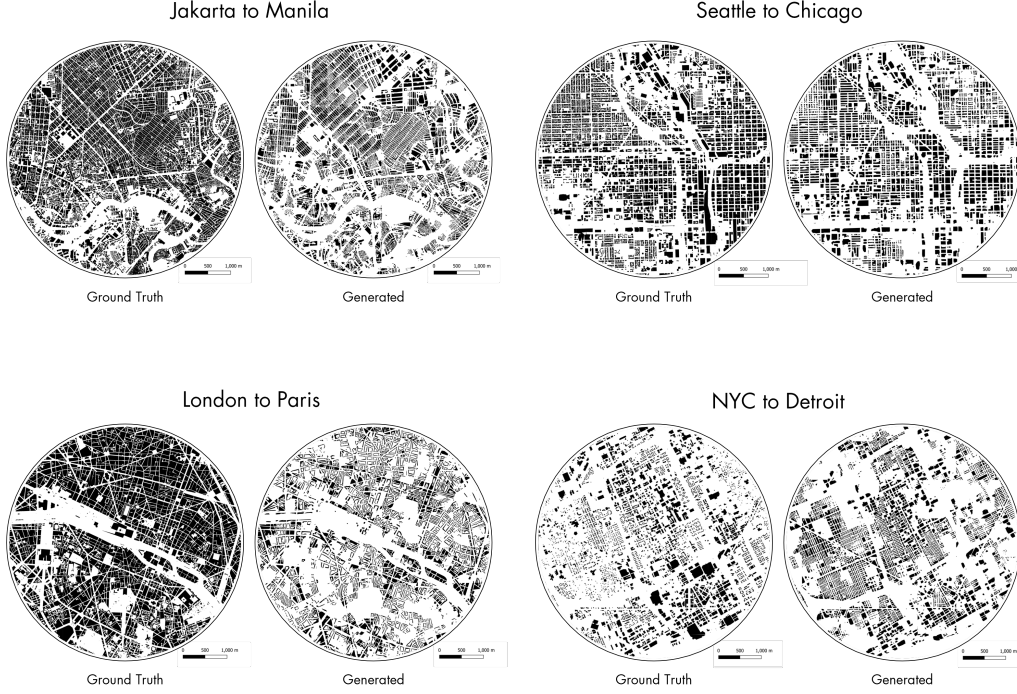
Figure 7: Selected results pairs comparing the generated cities with their ground truths.

the exact location of the generated footprints might not represent the ground truth, it can realistically represent the actual morphology in terms of average size, density and typology of buildings at the neighbourhood or city scale.

Globally, many cities suffer from the lack of building data (especially in OpenStreetMap, the leading global crowdsourced mapping dataset), and are thus excluded from a variety of studies where building and urban form data is an essential ingredient. In contrast, the same cities usually have full completeness of road network data available from OSM or other sources. Based on the trait of accurate generation of footprints with InstantCITY models, we posit that it is possible that a model trained in an area with 100% complete ground truth data can be used to check the data completeness of another area with similar typology based on solely street networks in the tested area. This is especially useful in cases where there is a drastic decrease in building data quality in different neighbourhoods or administrative regions in a metro area (i.e. heterogeneous data quality, which is typical for VGI sources). For example, in Jakarta, there are many cases where adjacent neighbourhoods can have considerably different levels of quality of data as shown in Figure 8. Similar phenomena also occur in

other locations such as Seattle, which is also shown in the same figure. In both cases, we observe that the urban pattern covers most of the tiles as seen in the satellite images, while a large number of buildings are missing in OSM.
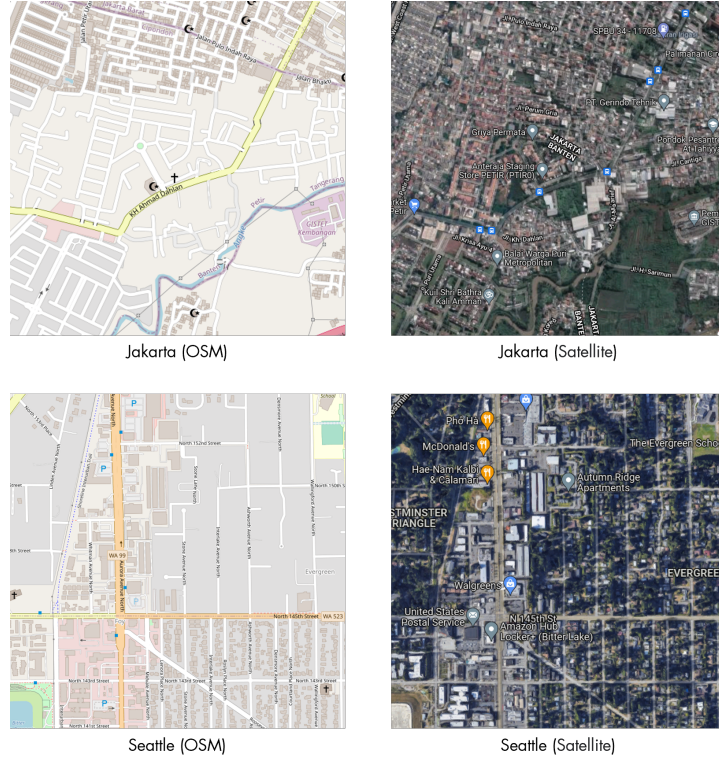

Jakarta (OSM)   Jakarta (Satellite)

Seattle (OSM)   Seattle (Satellite)

Figure 8: Drastic variation in building data quality (completeness) in the same city, which sometimes occurs at boundaries between metro areas and areas of activities of contributors. In both cases, we observe that the urban pattern covers most of the tile as seen in the satellite images while a large number of buildings are missing in the OSM representation. Our method is able to automatically assess the completeness at the neighbourhood level and identify undermapped areas, essentially unveiling a new approach for spatial data quality assessment. Source: (c) OpenStreetMap contributors and Google Maps.

With the state of the art of spatial data quality assessment methods, it is difficult to expose under-mapped areas without searching for them manually or comparing them to an authoritative dataset, a staple method in spatial data quality assessment. Thus, only a limited number of areas around the world can be assessed. Our model might be used to detect anomalies in the dataset by comparing the generated dataset to a region of interest, and it can be conducted at a large-scale regardless of jurisdictions and their authoritative data availability. Tiles within the region that are vastly different in terms of typology and density

can be flagged as incomplete for further manual action from the volunteers.

To validate the method, tiles from the ground truth of Seattle and Jakarta (For which we attested to as fully complete) are randomly selected and polygons within the tiles are randomly removed according to certain percentages, i.e. we are simulating intentional errors, a common approach in uncertainty propagation studies in GIS (Burnicki et al., 2007). This process creates an artificially under-sampled dataset where tiles that underwent random reduction can be labelled to verify the performance of the generative model.

Figure 9: Random reduction tile polygons in a city to simulate tiles that are undermapped. Prediction using generated dataset against the reduced dataset offers satisfactory accuracy (F1 Weighted Score: 0.924).

Figure 9 shows an example of the operation, removing certain percentages of

Table 4: Metric differences between the classes after random reduction (i.e. simulated varying degrees of completeness).

Class	PM ratio	Count ratio	Site Cover ratio	mIoU
Mapped	0.83	0.70	1.06	0.46
Partially Mapped	1.46	1.28	1.62	0.33
Unmapped	5.89	5.46	6.66	0.08

buildings from tiles, mirroring real-world scenarios of partial completeness. The process generates three classes:

1. Mapped – at least 80% mapped buildings.
2. Partially Mapped – at least 25% mapped buildings.
3. Unmapped – less than 25% mapped buildings.

Since the generated datasets represent the ground truth both in terms of site coverage and density, a tile that underwent a substantial reduction in polygons would have a large difference compared to the generated tile. Applying the comparison to the region of study, we can see from Table 4 that the mean metrics of tiles under difference classes differ significantly. The insight provided by the table may then be used to create thresholds to classify the quality of completion of the reduced dataset.

A simple heuristic threshold is applied to the Site Cover Ratio to classify tiles. Site Cover Ratio is calculated by dividing the site cover of the generated tile by the site cover of the target tile. A tile can be considered as 'Mapped' if the ratio is close or smaller than 1.2, indicating that the generated tile has similar coverage to the target tile. Similarly, a tile can be considered 'Partially Unmapped' if the ratio is larger than 1.2, indicating that the generated tile has at least 1.2 times the area than the target tile. Finally, a tile can be considered as 'Unmapped' if the ratio is larger than 4, indicating that the generated tile has at least 4 times the area compared to the target tile, potentially flagging severely undermapped areas and automatically tagging them as unreliable and requiring further attention.

Applying the above thresholds to both artificially reduced datasets of Seattle and Jakarta, the method achieved an F1 Weighted Score of 0.924 and the confusion matrix is shown in Figure 10. This is significant as the same threshold is applied to two cities with different morphologies, hinting at the potential universality of the model in other areas.
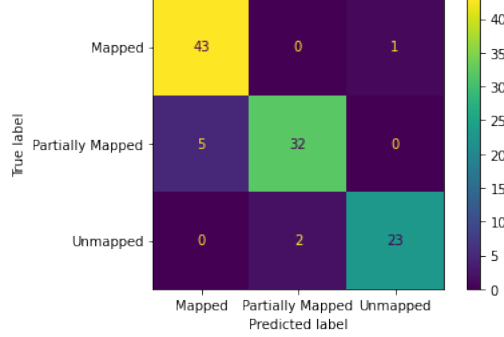
Figure 10: Confusion Matrix for the performance of automatically assessing the completeness of building data in Jakarta and Seattle using the novel approach we propose.

5. Discussions and limitations

From the experiments, we have shown that the InstantCITY model can generate high-quality synthetic data that captures the statistical structure of the original data. Translating into spatial sciences, the statistical structure of urban datasets are described and measured with urban morphology, such as urban texture, urban grain, and urban typologies.

The idea of Geographical Data Translation (GDT) could become an alternative way to model the growth of a city, check the quality of existing data, and transferring styles between different cities as postulated in Experiments 3 and 4. For example, InstantCITY models trained on different typologies can be used to quickly generate design options without building rules specifically for the desired morphology. This can be useful in the planning of new districts where planners could rapidly prototype a design by importing models trained on the desired urban forms to conduct a variety of simulations on the vector dataset. The same concept can also be applied to simulate future urban growth, assuming new regions of the city would continue from the existing urban patterns.

On the other hand, the high accuracy in data quality assessment explored in Experiment 4 can be adopted as a superior approach than rule-based algorithms in identifying areas with lower OSM data quality. The fast inference speed of GANs also makes the method scalable to large areas.

While certain limitations such as morphological inconsistencies between tiles and non-rectangular footprints exist in some cases as shown in the experiments, the presented concept is a pertinent area to focus in the future given the many asymmetries in different geospatial datasets. This concept is not limited to building footprints and road networks, but applicable to other geospatial data pairs

26

as long as there is a spatial similarity between the two datasets. For example, it might be possible to generate topologically correct secondary and tertiary street networks from primary road networks or to generate buildings and roads from geological features such as topography, water bodies, and forest boundaries.

Generally, GANs require a large amount of data during training and a common belief is that the more the data, the better the quality. However, as we have shown in the experiments, although the dataset at 500m/per tile has significantly more training samples that that at 1000m/per tile, the results are inferior. We have hypothesised that the cause of this phenomenon is due to a lack of contextual information at the 500m scale since the tiles might only capture a few buildings at this resolution. This meant that although the number of tiles have increased, the information contained in each tile became more fragmented. For example, larger buildings would be cut off and smaller buildings would lost their contextual relationships with adjacent roads and buildings. This phenomenon creates a dilemma in which one needs to find a balance between dataset size and the density of information in each tile. Although we found that models trained at 1000m/per tile are generally better, the exact optimised dataset scale for each city might vary as each would have different urban density. For example, fine-grained urban areas like Jakarta and Manilla with small and dense building footprints could use a higher zoom level (500m/tile), whereas coarse-grained areas such as Manhattan or Las Vegas needs a lower zoom level (1000m/tile or more) for the best prediction result.

Another challenge faced during training is the quality of the ground truth dataset. As shown in Experiment 4, ground truth building footprints may be entirely missing or have partial completeness in some areas. We have conducted a manual search for incomplete areas and tried our best in filtering out tiles with missing information. However, there is still a potential contamination of partially complete tiles in the training dataset. When applying the method to new cities, care needs to be taken to make sure the training data is comprised mostly of complete tiles. Otherwise, the model would learn the pattern of the incomplete dataset, resulting in lower output quality. Perhaps a way to circumvent this issue is to design a recursive elimination method using the completeness filter in Experiment 4. A few rounds of training and filtering by the model can be ran to filter out tiles that are undermapped.

6. Conclusion

Creating human faces, imagining street views, generating cities — GANs are opening up an exciting new era where AIs can now carry out tasks that were

previously only possible for humans.

Previous research has shown that GANs can translate, or 'create' realistic geospatial datasets by taking hints in a closely related dataset. This kind of visual inference was only possible for experienced urban planners, extracting the essences in a urban morphology and applying the nuances onto a new region without looking at the ground truth.

This paper pushes a new state-of-the-art for Geographical Data Translation. Our major contributions are methodological and the validation of the feasibility of GANs in generating geospatial data that can actually be used for spatial analyses rather than just giving a visual impression of synthetic content.

Compared to the previous state-of-the-art on raster data generation, the raster output from our model is 4 times higher in resolution. In the metrics for the vector output, the average error in Site Cover Percentage achieved -2.5% and a median building area as low as 8.6%. Applying the model for urban data quality control, the model achieved an F1 Weighted Score of 0.924 in a case study conducted on two cities with distinct morphologies.

While generating realistic synthetic datasets has been popular in other domains (Roth et al., 2020), in the domain of studying the urban form, it has not been so common, thus, we contribute with a new direction. Our work also introduces a new approach to automatically detect built-up areas missing from spatial datasets, especially volunteered instances such as OpenStreetMap. While simulating errors to understand uncertainty propagation has been common, to the extent of our knowledge, simulating varying completeness of spatial data has not been investigated before and is another contribution of ours.

We focused on buildings as they remain severely undermapped in many urban areas around the world. Our model could be utilised to generate urban forms in places where real labels are unavailable. These synthetic datasets could be sufficiently accurate for a variety of studies requiring data on the urban form.

We also hope that our paper could be one of the first steps in exploring how AIs can help designers and planners to make data driven decisions. The model's ability in translating learned morphologies onto other road networks could potentially remove the need for complex procedural modelling during design prototyping. This also leads to the discussion on how much AIs can take over designers' tasks in the future and how they would help designers to work more effectively.

In the future work, we plan to research the feasibility of the work to generate 2.5 or 3D building data. An increasing number of studies is taking advantage of the information of the height of buildings, which we did not cover in this work. Thanks to efforts measuring the heights of buildings at large-scale (Gui

and Qin, 2021; Recla and Schmitt, 2022; Pang and Biljecki, 2022), an increasing number of cities around the world has buildings mapped in 3D. We plan to take advantage of such data to investigate whether the method could be extended to generate basic 3D models that may add further value to use cases.

Acknowledgements

References

Abady, L., Barni, M., Garzelli, A., Tondi, B., 2020. GAN generation of synthetic multispectral satellite images, in: Bruzzone, L., Bovolo, F., Santi, E. (Eds.), Image and Signal Processing for Remote Sensing XXVI, International Society for Optics and Photonics. SPIE. pp. 122 – 133. URL: `https://doi.org/10.1117/12.2575765`, doi:`10.1117/12.2575765`.

Ahn, Y., Sohn, D.W., 2019. The effect of neighbourhood-level urban form on residential building energy use: A GIS-based model using building energy benchmarking data in seattle. Energy and Buildings 196, 124–133. URL: `https://doi.org/10.1016%2Fj.enbuild.2019.05.018`, doi:`10.1016/j.enbuild.2019.05.018`.

Andrade, H.J.A., Fernandes, B.J.T., 2020. Synthesis of satellite-like urban images from historical maps using conditional gan. IEEE Geoscience and Remote Sensing Letters , 1–4doi:`10.1109/LGRS.2020.3023170`.

Arribas-Bel, D., Fleischmann, M., 2022. Understanding (urban) spaces through form and function. Habitat International 128, 102641. doi:`10.1016/j.habitatint.2022.102641`.

Baier, G., Deschemps, A., Schmitt, M., Yokoya, N., 2021. Synthesizing optical and SAR imagery from land cover maps and auxiliary raster data. IEEE Transactions on Geoscience and Remote Sensing , 1–12doi:`10.1109/tgrs.2021.3068532`.

Balducci, F., 2019. Is OpenStreetMap a good source of information for cultural statistics? the case of italian museums. Environment and Planning B: Urban Analytics and City Science 48, 503–520. URL: `https://doi.org/10.1177%2F2399808319876949`, doi:`10.1177/2399808319876949`.

Barrington-Leigh, C., Millard-Ball, A., 2017. The world's open-source street map is more than 80% complete. PLoS One 12, 1–20.

Basiri, A., Haklay, M., Foody, G., Mooney, P., 2019. Crowdsourced geospatial data quality: challenges and future directions. International Journal of Geographical Information Science 33, 1588–1593. URL: `https://doi.org/10.1080%2F13658816.2019.1593422`, doi:`10.1080/13658816.2019.1593422`.

Beaulieu-Jones, B.K., Wu, Z.S., Williams, C., Lee, R., Bhavnani, S.P., Byrd, J.B., Greene, C.S., 2019. Privacy-preserving generative deep neural networks support clinical data sharing. Circulation: Cardiovascular Quality and Outcomes 12, e005122.

Biljecki, F., 2020. Exploration of open data in Southeast Asia to generate 3D building models. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences VI-4/W1-2020, 37–44. doi:10.5194/isprs-annals-vi-4-w1-2020-37-2020.

Biljecki, F., Chow, Y.S., 2022. Global Building Morphology Indicators. Computers, Environment and Urban Systems 95, 101809. doi:10.1016/j.compenvurbsys.2022.101809.

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

Birch, P.J., Browne, S.P., Jennings, V.J., Day, A.M., Arnold, D.B., 2001. Rapid procedural-modelling of architectural structures, in: Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage, Association for Computing Machinery, New York, NY, USA. pp. 187–196. URL: https://doi.org/10.1145/584993.585023, doi:10.1145/584993.585023.

Bowles, C., Gunn, R., Hammers, A., Rueckert, D., 2018. Gansfer learning: Combining labelled and unlabelled data for gan based data augmentation. arXiv preprint arXiv:1811.10669 .

Brock, A., Donahue, J., Simonyan, K., 2018. Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 .

Brovelli, M.A., Minghini, M., Molinari, M., Mooney, P., 2016. Towards an automated comparison of OpenStreetMap with authoritative road datasets. Transactions in GIS 21, 191–206. doi:10.1111/tgis.12182.

Burnicki, A.C., Brown, D.G., Goovaerts, P., 2007. Simulating error propagation in land-cover change analysis: The implications of temporal dependence. Computers, Environment and Urban Systems 31, 282–302. URL: https://doi.org/10.1016%2Fj.compenvurbsys.2006.07.005, doi:10.1016/j.compenvurbsys.2006.07.005.

Chen, W., Wu, A.N., Biljecki, F., 2021. Classification of urban morphology with deep learning: Application on urban vitality. Computers, Environment and Urban Systems 90, 101706. URL: https://www.sciencedirect.com/science/article/pii/S0198971521001137, doi:https://doi.org/10.1016/j.compenvurbsys.2021.101706.

Chen, X., Mishra, N., Rohaninejad, M., Abbeel, P., 2018. PixelSNAIL: An improved autoregressive generative model, in: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning, PMLR. pp. 864–872.

Christophe, S., Mermet, S., Laurent, M., Touya, G., 2022. Neural map style transfer exploration with GANs. International Journal of Cartography 8, 18–36. doi:10.1080/23729333.2022.2031554.

Engel, J., Agrawal, K.K., Chen, S., Gulrajani, I., Donahue, C., Roberts, A., 2019. Gansynth: Adversarial neural audio synthesis. arXiv preprint arXiv:1902.08710 .

Fink, T., Koenig, R., 2019. Integrated parametric urban design in grasshopper/rhinoceros 3d-demonstrated on a master plan in vienna .

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. The MIT Press.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. arXiv preprint arXiv:1406.2661 .

Grinberger, A.Y., Schott, M., Raifer, M., Zipf, A., 2021. An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. Transactions in GIS 25,

622–641. URL: `https://doi.org/10.1111%2Ftgis.12746`, doi:`10.1111/tgis.12746`.

Groenewegen, S.A., Smelik, R.M., de Kraker, K.J., Bidarra, R., 2009. Procedural city layout generation based on urban land use models. Short Paper Proceedings of Eurographics 2009 .

Gui, S., Qin, R., 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. ISPRS Journal of Photogrammetry and Remote Sensing 181, 1–19. URL: `https://doi.org/10.1016%2Fj.isprsjprs.2021.08.025`, doi:`10.1016/j.isprsjprs.2021.08.025`.

Han, Z.Y., Wang, J., Fan, H., Wang, L., Zhang, P., 2018. Unsupervised generative modeling using matrix product states. Physical Review X 8, 031012.

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

Heikinheimo, V., Tenkanen, H., Bergroth, C., Järv, O., Hiippala, T., Toivonen, T., 2020. Understanding the use of urban green spaces from user-generated geographic information. Landscape and Urban Planning 201, 103845. doi:`10.1016/j.landurbplan.2020.103845`.

Heris, M.P., Foks, N.L., Bagstad, K.J., Troy, A., Ancona, Z.H., 2020. A rasterized building footprint dataset for the united states. Scientific Data 7. URL: `https://doi.org/10.1038%2Fs41597-020-0542-3`, doi:`10.1038/s41597-020-0542-3`.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems 30.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. science 313, 504–507.

Hou, Y., Biljecki, F., 2022. A comprehensive framework for evaluating the quality of street view imagery. International Journal of Applied Earth Observation and Geoinformation 115, 103094. doi:`10.1016/j.jag.2022.103094`.

Huang, B., Wang, J., 2020. Big spatial data for urban and environmental sustainability. Geospatial Information Science 23, 125–140.

Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. CVPR .

Jacobs, K.T., Mitchell, S.W., 2020. OpenStreetMap quality assessment using unsupervised machine learning methods. Transactions in GIS 24, 1280–1298. URL: `https://doi.org/10.1111%2Ftgis.12680`, doi:`10.1111/tgis.12680`.

Jiang, K., Wang, Z., Yi, P., Wang, G., Lu, T., Jiang, J., 2019. Edge-enhanced gan for remote sensing image superresolution. IEEE Transactions on Geoscience and Remote Sensing 57, 5799–5812. doi:`10.1109/TGRS.2019.2902431`.

Jin, M., Sun, R., Yang, X., Yan, M., Chen, L., 2022. Remote sensing-based morphological analysis of core city growth across the globe. Cities , 103982doi:`10.1016/j.cities.2022.103982`.

Kang, Y., Gao, S., Roth, R.E., 2019. Transferring multiscale map styles using generative adversarial networks. International Journal of Cartography 5, 1–27. doi:`10.1080/23729333.2019.1615729`, arXiv:`1905.02200`.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T., 2021. Alias-free generative adversarial networks, in: Proc. NeurIPS.

Kim, J.S., Kavak, H., Crooks, A., 2018. Procedural city generation beyond game development. SIGSPATIAL Special 10, 34–41.

Lee, J.G., Kang, M., 2015. Geospatial big data: challenges and opportunities. Big Data Research 2, 74–81.

Leonard, A., Wheeler, S., McCulloch, M., 2022. Power to the people: Applying citizen science and computer vision to home mapping for rural energy access. International Journal of Applied Earth Observation and Geoinformation 108, 102748. doi:10.1016/j.jag.2022.102748.

Li, H., Herfort, B., Huang, W., Zia, M., Zipf, A., 2020a. Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in mozambique. ISPRS Journal of Photogrammetry and Remote Sensing 166, 41–51. URL: https://doi.org/10.1016%2Fj.isprsjprs.2020.05.007, doi:10.1016/j.isprsjprs.2020.05.007.

Li, H., Herfort, B., Lautenbach, S., Chen, J., Zipf, A., 2022a. Improving OpenStreetMap missing building detection using few-shot transfer learning in sub-Saharan Africa. Transactions in GIS doi:10.1111/tgis.12941.

Li, J., Chen, Z., Zhao, X., Shao, L., 2020b. Mapgan: An intelligent generation model for network tile maps. Sensors 20, 3119.

Li, J., Wu, Z., Hu, Z., Zhang, J., Li, M., Mo, L., Molinier, M., 2020c. Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. ISPRS Journal of Photogrammetry and Remote Sensing 166, 373–389. URL: https://www.sciencedirect.com/science/article/pii/S0924271620301787, doi:https://doi.org/10.1016/j.isprsjprs.2020.06.021.

Li, M., Koks, E., Taubenböck, H., Vliet, J.v., 2020d. Continental-scale mapping and analysis of 3D building structure. Remote Sensing of Environment 245, 111859. doi:10.1016/j.rse.2020.111859.

Li, M., Wang, Y., Rosier, J.F., Verburg, P.H., Vliet, J.v., 2022b. Global maps of 3D built-up patterns for urban morphological analysis. International Journal of Applied Earth Observation and Geoinformation 114, 103048. doi:10.1016/j.jag.2022.103048.

Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., et al., 2016. Geospatial big data handling theory and methods: A review and research challenges. ISPRS journal of Photogrammetry and Remote Sensing 115, 119–133.

Li, Y., Liu, S., Yang, J., Yang, M.H., 2017. Generative face completion, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3911–3919.

Lipson, M.J., Nazarian, N., Hart, M.A., Nice, K.A., Conroy, B., 2022. A Transformation in City-Descriptive Input Data for Urban Climate Models. Frontiers in Environmental Science 10, 866398. doi:10.3389/fenvs.2022.866398.

Litjens, G., Ciompi, F., Wolterink, J.M., de Vos, B.D., Leiner, T., Teuwen, J., Išgum, I., 2019. State-of-the-art deep learning in cardiovascular image analysis. JACC: Cardiovascular Imaging 12, 1549–1565. URL: https://doi.org/10.1016%2Fj.jcmg.2019.06.009, doi:10.1016/j.jcmg.2019.06.009.

Luo, J., Zhao, T., Cao, L., Biljecki, F., 2022. Semantic Riverscapes: Perception and evaluation of linear landscapes from oblique imagery using computer vision. Landscape and Urban Planning 228, 104569. doi:10.1016/j.landurbplan.2022.104569.

Ma, J., Yu, W., Chen, C., Liang, P., Guo, X., Jiang, J., 2020. Pan-gan: An unsupervised pan-sharpening method for remote sensing image fusion. Information Fusion 62, 110–120. URL: https://www.sciencedirect.com/science/article/pii/S1566253520302591, doi:https://doi.org/10.1016/j.inffus.2020.04.006.

Maaløe, L., Sønderby, C.K., Sønderby, S.K., Winther, O., 2016. Auxiliary deep generative models, in: International conference on machine learning, PMLR. pp. 1445–1453.

Majic, I., Naghizade, E., Winter, S., Tomko, M., 2021. There is no way! Ternary qualitative spatial reasoning for error detection in map data. Transactions in GIS 25, 2048–2073. doi:`10.1111/tgis.12765`.

Milojevic-Dupont, N., Hans, N., Kaack, L.H., Zumwald, M., Andrieux, F., Soares, D.d.B., Lohrey, S., Pichler, P.P., Creutzig, F., 2020. Learning from urban form to predict building heights. PLOS ONE 15, e0242010. doi:`10.1371/journal.pone.0242010`.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 .

Ng, V., Hofmann, D., 2018. Scalable feature extraction with aerial and satellite imagery, pp. 145–151. doi:`10.25080/Majora-4af1f417-015`.

Pang, H.E., Biljecki, F., 2022. 3D building reconstruction from single street view images using deep learning. International Journal of Applied Earth Observation and Geoinformation 112, 102859. doi:`10.1016/j.jag.2022.102859`.

Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y., 2019. Semantic image synthesis with spatially-adaptive normalization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2337–2346.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library, in: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., 2016. Context encoders: Feature learning by inpainting, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2536–2544.

Quan, S.J., 2022. Urban-GAN: An artificial intelligence-aided computation system for plural urban design. Environment and Planning B: Urban Analytics and City Science , 239980832211005doi:`10.1177/23998083221100550`.

Quintana, M., Schiavon, S., Tham, K.W., Miller, C., 2020. Balancing thermal comfort datasets. Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation , 120–129doi:`10.1145/3408308.3427612`, arXiv:`2009.13154`.

Rachele, J.N., Wang, J., Wijnands, J.S., Zhao, H., Bentley, R., Stevenson, M., 2021. Using machine learning to examine associations between the built environment and physical function: A feasibility study. Health & Place 70, 102601. doi:`10.1016/j.healthplace.2021.102601`.

Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 .

Recla, M., Schmitt, M., 2022. Deep-learning-based single-image height reconstruction from very-high-resolution SAR intensity data. ISPRS Journal of Photogrammetry and Remote Sensing 183, 496–509. URL: `https://doi.org/10.1016%2Fj.isprsjprs.2021.11.`

012, doi:10.1016/j.isprsjprs.2021.11.012.

Roth, J., Martin, A., Miller, C., Jain, R.K., 2020. SynCity: Using open data to create a synthetic city of hourly building energy estimates by integrating data-driven and physics-based methods. Applied Energy 280, 115981. URL: https://doi.org/10.1016%2Fj.apenergy.2020.115981, doi:10.1016/j.apenergy.2020.115981.

Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M.M., 2016. A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science 31, 139–167. URL: https://doi.org/10.1080%2F13658816.2016.1189556, doi:10.1080/13658816.2016.1189556.

Smelik, R.M., Tutenel, T., Bidarra, R., Benes, B., 2014. A survey on procedural modelling for virtual worlds, in: Computer Graphics Forum, Wiley Online Library. pp. 31–50.

Songchon, C., Wright, G., Beevers, L., 2021. Quality assessment of crowdsourced social media data for urban flood management. Computers, Environment and Urban Systems 90, 101690. URL: https://doi.org/10.1016%2Fj.compenvurbsys.2021.101690, doi:10.1016/j.compenvurbsys.2021.101690.

Sun, Y., Montazeri, S., Wang, Y., Zhu, X.X., 2020. Automatic registration of a single SAR image and GIS building footprints in a large-scale urban area. ISPRS Journal of Photogrammetry and Remote Sensing 170, 1–14. doi:10.1016/j.isprsjprs.2020.09.016.

Sundaram, R.C., Naghizade, E., Borovica-Gajic, R., Tomko, M., 2021. Can you fixme? an intrinsic classification of contributor-identified spatial data issues using topic models. International Journal of Geographical Information Science 36, 1–30. URL: https://doi.org/10.1080%2F13658816.2021.1893323, doi:10.1080/13658816.2021.1893323.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

Tobiáš, P., Cajthaml, J., 2020. Models of cultural heritage buildings in a procedurally generated geospatial environment. Transactions in GIS 25, 1104–1122. URL: https://doi.org/10.1111%2Ftgis.12727, doi:10.1111/tgis.12727.

Toker, A., Zhou, Q., Maximov, M., Leal-Taixé, L., 2021. Coming down to earth: Satellite-to-street view synthesis for geo-localization. arXiv:2103.06818.

Wagner, F., Milojevic-Dupont, N., Franken, L., Zekar, A., Thies, B., Koch, N., Creutzig, F., 2022. Using explainable machine learning to understand how urban form shapes sustainable mobility. Transportation Research Part D: Transport and Environment 111, 103442. doi:10.1016/j.trd.2022.103442.

Wang, M., Chen, Z., Rong, H.H., Mu, L., Zhu, P., Shi, Z., 2022. Ridesharing accessibility from the human eye: Spatial modeling of built environment with street-level images. Computers, Environment and Urban Systems 97, 101858. doi:10.1016/j.compenvurbsys.2022.101858.

Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B., 2018. High-resolution image synthesis and semantic manipulation with conditional gans, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Wu, A.N., Biljecki, F., 2021. Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. Landscape and Urban Planning 214, 104167. doi:10.1016/j.landurbplan.2021.104167.

Wu, A.N., Biljecki, F., 2022. GANmapper: geographical data translation. International Journal of Geographical Information Science 36, 1394–1422. doi:10.1080/13658816.2022.

2041643.

Wu, A.N., Stouffs, R., Biljecki, F., 2022. Generative Adversarial Networks in the built environment: A comprehensive review of the application of GANs across data types and scales. Building and Environment , 109477doi:10.1016/j.buildenv.2022.109477.

Yan, K., Chong, A., Mo, Y., 2020a. Generative adversarial network for fault detection diagnosis of chillers. Building and Environment 172, 106698. doi:10.1016/j.buildenv.2020.106698.

Yan, Y., Feng, C.C., Huang, W., Fan, H., Wang, Y.C., Zipf, A., 2020b. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. International Journal of Geographical Information Science 34, 1–27. doi:10.1080/13658816.2020.1730848.

Yan, Y., Huang, B., 2022. Estimation of building height using a single street view image via deep neural networks. ISPRS Journal of Photogrammetry and Remote Sensing 192, 83–98. doi:10.1016/j.isprsjprs.2022.08.006.

Yeboah, G., de Albuquerque, J.P., Troilo, R., Tregonning, G., Perera, S., Ahmed, S.A.K.S., Ajisola, M., Alam, O., Aujla, N., Azam, S.I., Azeem, K., Bakibinga, P., Chen, Y.F., Choudhury, N.N., Diggle, P.J., Fayehun, O., Gill, P., Griffiths, F., Harris, B., Iqbal, R., Kabaria, C., Ziraba, A.K., Khan, A.Z., Kibe, P., Kisia, L., Kyobutungi, C., Lilford, R.J., Madan, J.J., Mbaya, N., Mberu, B., Mohamed, S.F., Muir, H., Nazish, A., Njeri, A., Odubanjo, O., Omigbodun, A., Osuh, M.E., Owoaje, E., Oyebode, O., Pitidis, V., Rahman, O., Rizvi, N., Sartori, J., Smith, S., Taiwo, O.J., Ulbrich, P., Uthman, O.A., Watson, S.I., Wilson, R., Yusuf, R., 2021. Analysis of OpenStreetMap data quality at different stages of a participatory mapping process: Evidence from slums in africa and asia. ISPRS International Journal of Geo-Information 10, 265. URL: https://doi.org/10.3390%2Fijgi10040265, doi:10.3390/ijgi10040265.

Yeh, R.A., Chen, C., Yian Lim, T., Schwing, A.G., Hasegawa-Johnson, M., Do, M.N., 2017. Semantic image inpainting with deep generative models, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5485–5493.

Yoon, J., Jarrett, D., van der Schaar, M., 2019. Time-series generative adversarial networks, in: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf.

Yu, Y., Srivastava, A., Canales, S., 2021. Conditional lstm-gan for melody generation from lyrics. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 17, 1–20.

Zakharov, E., Shysheya, A., Burkov, E., Lempitsky, V., 2019. Few-shot adversarial learning of realistic neural talking head models, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9459–9468.

Zhang, J., Fukuda, T., Yabuki, N., 2021a. Automatic object removal with obstructed façades completion using semantic segmentation and generative adversarial inpainting. IEEE Access 9, 117486–117495. doi:10.1109/ACCESS.2021.3106124.

Zhang, J., Xu, L., Shabunko, V., Tay, S.E.R., Sun, H., Lau, S.S.Y., Reindl, T., 2019. Impact of urban block typology on building solar potential and energy use efficiency in tropical high-density city. Applied Energy 240, 513–533. URL: https://doi.org/10.1016%2Fj.apenergy.2019.02.033, doi:10.1016/j.apenergy.2019.02.033.

Zhang, Z., Zhou, Q., Xu, Y., Ma, L., Iwasaki, A., 2021b. Remote sensing image jitter restoration based on deep generative adversarial network, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE. URL: `https://doi.org/10.1109%2Figarss47720.2021.9554491`, doi:10.1109/igarss47720.2021.9554491.

Zhao, B., Zhang, S., Xu, C., Sun, Y., Deng, C., 2021. Deep fake geography? when geospatial data encounter artificial intelligence. Cartography and Geographic Information Science 48, 338–352. doi:10.1080/15230406.2021.1910075.

Zhou, Q., 2017. Exploring the relationship between density and completeness of urban building data in OpenStreetMap for quality estimation. International Journal of Geographical Information Science 32, 257–281. URL: `https://doi.org/10.1080%2F13658816.2017.1395883`, doi:10.1080/13658816.2017.1395883.

Zhu, F., Ye, F., Fu, Y., Liu, Q., Shen, B., 2019. Electrocardiogram generation with a bidirectional lstm-cnn generative adversarial network. Scientific reports 9, 1–11.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.