Towards a Deep Automatic Generation of Figure-ground Maps

Lukas Arzoumanidis, Jonathan Hecht, Youness Dehbi

Computational Methods Lab, HafenCity University, Hamburg, Germany - (lukas.arzoumanidis, jonathan.hecht, youness.dehbi)
@hcu-hamburg.de

Keywords: Generative Adversarial Networks, Geographical Data Translation, Figure-ground Maps, Urban Morphology, Built Density, Volunteered Geographic Information.

Abstract

Figure-ground maps play a key role in many disciplines where urban planning or analysis is involved. In this context, the automatic generation of such maps with respect to certain requirements and constraints is an important task. This paper presents a first step towards a deep automatic generation of figure-ground maps where the built density of the generated scenes is controlled and taken into account. This is preformed building upon a Geographic Data Translation model which has been applied to generate less available geospatial features, e.g. building footprints, from more widely available geospatial data, e.g. street network data, using conditional Generative Adversarial Networks. A novel processing approach is introduced to incorporate the population density and the built density accordingly. Furthermore, the impact of both the level of detail of the street network, i.e. its sparsity or density, and the spatial resolution of the training data on the generated figure-ground maps has been investigated. The generated maps and the qualitative results reveal an obvious impact of these parameters on the layout of built and unbuilt areas. Our approach paves the way for the expansion of existing districts by figure-ground maps of future neighbourhoods considering factors such as density and further parameters which will be subject of future work.

1. Introduction

To counteract the growth of sealed surfaces and simultaneously create new living space in cities, urban planning plays a vital role. Current paradigms in urban planning suggest to rebuild or upgrade existing parcels or buildings instead of creating new districts on undeveloped land. In such case, a construction company redevelops a parcel with a goal, e.g. to maximize revenue or to build the most possible flats while respecting certain constraints, e.g. the building law or the physical dimensions of the parcel. In an initial planning phase, a draft representing the broad layout of the street network and the arrangement and geometry of the building footprints is laid out. This initial drafting process lays the foundation of the following planning phases, where each phase facilitates the draft further with detail to eventually become a site plan for a development project. Usually, the creation of such drafts builds upon the so called figure-ground maps as those visualize the existing built structures while demonstrating the relationship between built and unbuilt spaces in cities. In such maps, buildings are depicted as black solid mass (figure) while streets or open spaces are represented as white void (ground) (Wang et al., 2024).

In practice, it is rarely sufficient to create a single draft for a urban development project. Instead, several hundreds of initial drafts might be needed to visualize different and alternative design ideas. However, drafting is often a tedious and time consuming process as it involves mainly analogous or digital manual drawing or crafting. Depending on the scale and specifications of the parcel and the experience level of the planner, such a task could be overwhelming. Besides, there are many further requirements and constraints to consider such as the desired level of building density, the construction law or the physical dimensions of the plot. To support planners with this endeavour, this paper proposes a first approach towards a deep automatic generation of figure-ground maps with the possibility to control the built density of urban structures. This approach paves the way for the expansion of existing districts by figure-ground maps of future neighbourhoods considering influencing factors such as population and built density.

One recent research field that could pave the way towards the automatic generation of figure-ground maps is Geographic Data Translation (GDT) which follows the idea of generating less abundant geospatial datasets by learning associations from other more plentiful datasets. In this context, two new approaches *GANmapper* and *InstantCity* which use conditional Generative Adversarial Networks (cGAN) to tackle GDT have been proposed (Wu and Biljecki, 2023, 2022). These approaches capitalize on more commonly found geospatial features, e.g. street networks, to generate less common features such as building footprints by leveraging on their mutual relationship.

This paper investigates whether the recently proposed approach of Wu and Biljecki (2023) is able to capture and reflect higher and lower built densities of urban structures. By leveraging a latent relationship between the population and the built density, we present a novel training data processing that precisely captures the desired level of built density which will be explained in more detail in Section 3.2. Hence, the main contribution of this paper consists in the incorporation of new factors into the automatic generation of figure-ground maps. In particular, we focused on the assessment of the impact of the level of detail of the underlying street network, the spatial resolutions and the built density.

The remainder of this paper is as follows. Before going into the details of the methodology of our figure-ground map generation in Section 3, we give a review of related work in Section 2. The setup of our experiments and their findings are presented in Section 4. Section 5 concludes the paper and gives outlooks for future research.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W5-2024-33-2024 | © Author(s) 2024. CC BY 4.0 License.

2. Related Work

The analysis of the surrounding urban morphology is fundamental for new building projects. According to Wang et al. (2024) urban morphology studies heavily rely on morphometrics such as building footprints or street lengths. In this context, they developed a method for learning morphology features based on figure-ground maps where they compare urban form types in a fully unsupervised manner. In their approach, they apply a visual representation learning model called *SimCLR* to capture the layout of building groups (Wang et al., 2024). However, several factors may influence the underlying morphology. In this context, Hijazi et al. (2017) engineered a GIS-based approach to quantify, for instance, the homogeneity of urban structures as an important factor by extracting attributes such as the angles or distances between buildings directly from building footprints. In order to incorporate such influencing factors, our approach builds upon ideas of Wu and Biljecki (2023, 2022). These approaches have been dedicated to create geospatial data using conditional Generative Adversarial Networks following the image-to-image translation paradigm. In their work, street network data is used as input to generate building footprints with the goal to support applications in the fields of urban morphology analysis and urban simulations which often suffer from missing spatial details.

Recently, an approach developing a GAN-based end-to-end generative model for the 2D and 3D building layout generation has been proposed (Jiang et al., 2023b). They stated that most approaches overlook the impact of site attributes on built structures, hampering their potential for further evaluation and informed decision making. The authors conducted experiments with conditional vectors showing improved performance in different scenarios. Similarly, Jiang et al. (2024) proposed an approach for automating site planning with integrated domain knowledge of the built environment with the goal to improve context-awareness. Their developed Generative Adversarial Network called *CAIN-GAN* is supposed to not only be capable of synthesized visually realistic and semantically reasonable design solutions but also useful for urban sustainability simulations.

Many researchers applied formal grammars, particularly shape grammars (Stiny, 1980), as another paradigm to generate real-world man-made objects. For instance, Gong et al. (2020) generated urban fabric in orthogonal and non-orthogonal urban landscapes based on a pre-defined set of shape rules. However, these approaches are in general suffering under the overhead of the manual design of the grammar rules by experts. For more details on the topic of generative urban design, the interested reader is referred to Jiang et al. (2023a).

3. Methodology

The following sections will introduce the acquisition and processing of training data and describe the applied model for the generation of building footprints based on different street types building upon the ideas of Geographic Data Translation presented by Wu and Biljecki (2022) and further taking the population and street network density into account.

3.1 Model architecture

The network architecture of the applied model is a type of Image-to-Image conditional GAN (Isola et al., 2017) that can translate

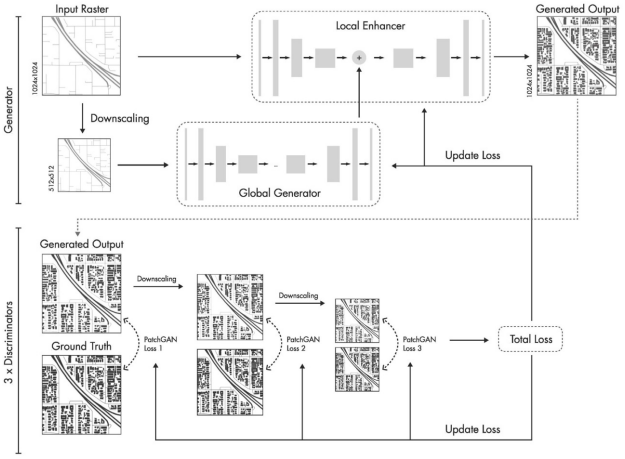

Figure 1. *InstantCITY* architecture which our approach builds upon (Wu and Biljecki, 2023).

input image data such as street networks to an output image facilitated with generated building footprints (Wu and Biljecki, 2023). Using different resolutions, the underlying network is trained on two residual networks namely a Global Generator and a Local Enhancer. In each forward pass, the generator tries to create outputs that could 'trick' the discriminator into thinking the outputs are 'real' while the discriminator will learn to classify the outputs as 'fake' and the ground truth as 'real' (Wu and Biljecki, 2023). Discriminators in the model compare the encoded features form both the generated images and the ground truth to indicate whether the generated image is real or fake. At the end of each forward pass, the loss of the generator and the discriminator is evaluated and their weights are updated accordingly until the generator's outputs are compelling enough so that the discriminator detects half of them as real images (Wu and Biljecki, 2023). Figure 1 shows an overview on the network architecture of InstantCITY which builds the basis of our paper taking further constraints into consideration.

3.2 Processing

To generate figure-ground maps, the previously described model is provided with training data consisting of two different sets of image pairs. Each image pair consists of an input and an output image where the model is supposed to learn the visual patterns to transform an input image to an output image accordingly. An input image resembles the street network with dif-

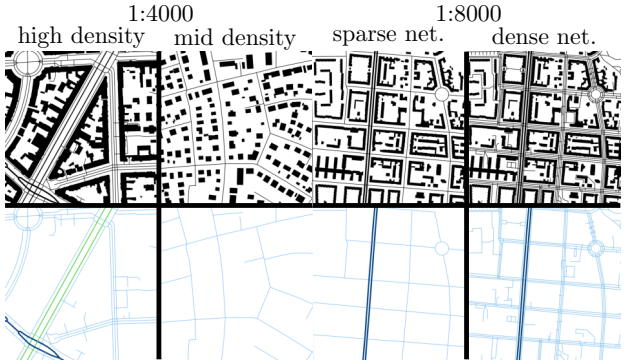

Figure 2. Example of training image pairs for different datasets showing spatial resolutions of 1:4000 vs. 1:8000, sparse vs. dense street network and high vs. mid built density.

color	line width (in px)	type
#ffb301	12 & 9	trunk, trunk_link
#840000	7.5 & 6	motorway, motorway_link
#ff1a01	12 & 6	primary, primary_link
#014182	9 & 3	secondary, secondary_link
#58d751	6 & 3	tertiary, tertiary_link
#75bbfd	3	access roads

Table 1. CRHD configuration applied in this work.

ferent street types indicated by different line-width and coloring accordingly. The output image represents the figure-ground map showing the street network including building footprints. Based on the road network and the building footprints that are extracted from OpenStreetMap (OSM), we can generate an arbitrary amount of such training image pairs. As mentioned, the geometry of the street network has to represent the street type and therefore is converted into a Coloured Road Hierarchy Diagram (CRHD) as proposed by (Wu and Biljecki, 2023). A CRHD differentiates street types by their assigned line-width and coloring as can be seen in Table 1 and Figure 2 (Chen et al., 2021). Ground truth images are rendered based on the building footprint geometry as well as the street network geometry where every geometry is colored in solid black and the street network geometry which has the same line-width as defined for the CRHD used in the input images. The types of roads and paths which have been selected are also listed in Table 1. As they belong to the same level of hierarchy, the types residential, service, living_street, footway, path, pedestrian and unclassified are grouped into *access roads*. Each hierarchy has been associated to a pre-defined color. For the sake of replicability, the used colors and the line widths are also listed in the same table.

The European Commission and the Federal Ministry of Transport and Digital Infrastructure (Bundesministerium für Verkehr und digitale Infrastruktur) divided the population density for Germany into three different categories for contiguous grid cells of 1 km^2 as highlighted in Figure 3 European Commission (2016); BMVI (2018). A $high$-density grid cell (city) corresponds to a population density of at least 1.500 inhabitants per km^2 with a minimum total population of 50.000. A mid-density grid cell (suburb) refers to a population density of at least 300 inhabitants per km^2 with a minimum total population of 5.000. A low-density grid cell has a population density below 300 inhabitants per km^2. In this work, these categories will be applied as a template for extracting the training data from OSM according to their population density as highlighted in Figure 4. As performed for the city of Hamburg, we overlay the population density and extract training images according to their spatial boundary for the ten biggest cities in Germany according to their total population. In the following, the population density based extraction of training images summarized in Figure 4 will be explained in more detail.

To render input and output image pairs, the *Atlas*-tool available in QGIS has been used. This tool allows for subsequent map rendering to automatically extract the training images for a specific spatial extent, scale and density category as can be seen in both Figure 2 and Figure 4. Subsequently, the training images for each population density are rendered in a resolution of 1024×1024 pixel and stored as a PNG file to match the input size and file type requirements of the underlying model designed by Wu and Biljecki (2023).

In total, we generate different training datasets with the following characteristics:

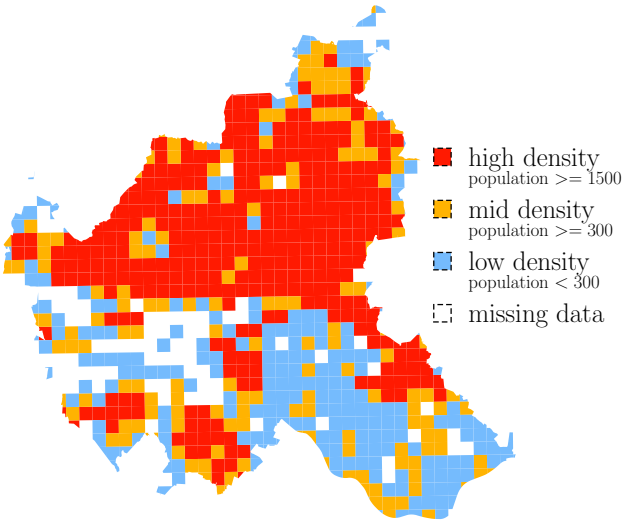

Figure 3. Population density grid cells representing defined categories for the city of Hamburg, Germany. Colors distinguish the different population densities. The categories have been derived according to the definition of the European Commission.

- Spatial resolution: 1:4000 or 1:8000
- Built density: high or mid
- Street network: sparse or dense

For each training dataset, we extracted the image pairs for a scale of 1:4000 and 1:8000 in order to provide high details to individual buildings or a neighborhood of building blocks as highlighted in Figure 2. Additionally, according to Wu and Biljecki (2023) at these scales the model is supposed to be able to generate artificial images with sharper corners that would represent real images more closely. The added value of this paper, is the consideration of the density of the street network and, hence, the investigation of its potential impact on the trained model and the according generated figure-ground maps. Therefore, we created two different training datasets. The first one does not include *access roads* and, hence, represents sparse street network data, whereas the second comprises such roads and, thus, reflects dense street network data as can be seen in Figure 2. Further, the population and the according built density has been taken into account in order to generate figure-ground maps depending on this parameter. Thus, the model has been trained on two different high and mid built densities as already depicted in Figure 2.

In general, categorizing the training data after the above mentioned levels of population densities results in the expected outcomes. For regions with high population densities, we can observe a high built density represented through closed building development. The latter is characterized by a continuous and cohesive built structure. Similarly, mid built densities show the expected open building development typically expressed by a more spacious and less densely constructed built structure. However, this pattern is sometimes violated contrasting the underlying training data.

To overcome this problem, we merged both the high and the mid built density training data resulting in two datasets for sparse and dense street networks. Analogously, two datasets with different spatial resolutions, namely 1:4000 and 1:8000 have been

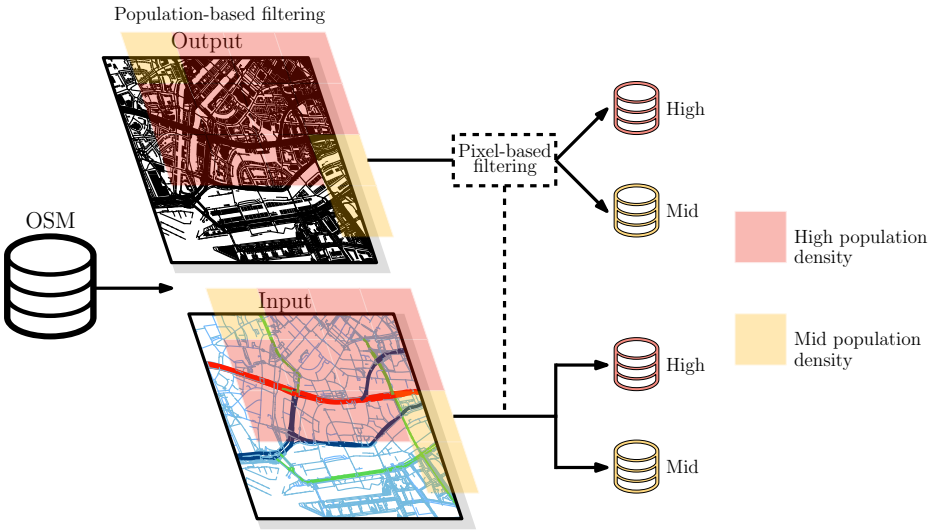

Figure 4. Illustration of our training data processing pipeline. High, mid and low population density data is provided as 1 km^2 grid cells by Eurostat but only high and mid population density data is processed.

generated. To further refine the four datasets and hence reflect well the level of density, we apply a *pixel-based filtering*. Herewith, image pairs are sorted according to their proportion of unbuilt area. An image comprising more than 82% white pixels is then removed from the training dataset. Images with less than 72% white pixels have been assigned to the training dataset for high built densities whereas those having a proportion of white pixels between 72% and 82% have been associated to the training dataset with mid built density accordingly. The thresholds used to refine our training datasets have been determined based on a qualitative visual inspection of the underlying images. Since the levels of built density might differ from country to country, the thresholds might be re-evaluated for cities outside of Germany. In German cities, a low population density can be predominantly stated in industrial districts such as the south of Hamburg as depicted in Figure 3. In this paper, we focused on districts with mid and high density and omitted those corresponding to low density. For each training dataset, the image pairs are split into training and testing data with a proportion of 90% and 10% respectively.

4. Experimental Results

The following section describes the training data in more details and gives insight into the achieved experimental results which have been evaluated by qualitative visual inspection and quantitative metrics.

4.1 Datasets

To produce our training data, we use OSM[1] and population density data[2]. In order to reflect the urban morphology of German cities, we focus on the ten biggest cities for training and generating figure-ground maps. This allows for addressing different population densities as important factor influencing the design of such diagrams. The following ten most populated cities in Germany have been selected accordingly: Berlin, Hamburg, Munich, Cologne, Frankfurt, Stuttgart, Düsseldorf, Leipzig, Dortmund, Essen (Statista, 2024). After downloading the OSM data of each city, the street network data and the building footprints are cropped to the administrative boundaries of each respective city. Analogously, the population density data is structured in contiguous grid cells of size 1 km^2 and matched to the administrative boundaries of the aforementioned ten biggest cities in Germany.

[1] https://download.geofabrik.de/europe/germany.html
[2] https://ec.europa.eu/eurostat/de/web/gisco/geodata/reference-data/population-distribution-demography/geostat

4.2 Configuration & Accuracy Measures

We applied the model proposed by Wu and Biljecki (2023) following the advised hyperparameters which yielded the best results in their experiments. The model was implemented in PyTorch and CUDA accelerator. In this context, our experiments have been conducted with PyTorch version 2.2.0 with CUDA version 12.3. While training, the model consumed roughly 11 GB of VRAM. A starting learning rate of 0.0002 and a batch size of 1 have been chosen with a training time amounting 100 epochs for each experiment.

One of the most commonly applied measure to assess the performance of a GAN is the Fréchet Inception Distance (FID) proposed by Heusel et al. (2017). The FID is a distance metric which captures the similarity of images created by generative models to real images and thereby expresses the models overall quality. Formally, with FID we measure the difference between two Gaussian distributions using the Fréchet distance (Har-Peled and Raichel, 2014). In this work, the first distribution corresponds to the dataset of images generated by our approach, while the other distribution corresponds to the reference dataset of images containing the ground truth, i.e. figure-ground maps of identical spatial location. Each parametric Gaussian distribution is defined by it's mean μ and covariance matrix Σ which have been derived after transforming the real and generated images from the image space to latent vector embeddings using the final coding layer of an inception model as can be seen in Equation 1. This way, the FID compares features that correspond to real-world objects rather than directly comparing generated and real images pixel-wise. As a result, a higher FID indicates a larger difference between the generated and the real image while a smaller FID indicates a higher similarity.

In our case, this is in general optically reflected by more correct representations of shapes, sizes, and densities of generated building footprints.

$$FID = ||\mu_X - \mu_Y||^2 - Tr(\Sigma_X + \Sigma_Y - 2\sqrt{\Sigma_X \Sigma_Y}) \quad (1)$$

4.3 Interpretation & Discussion

Figure 5. Comparison of two generated figure-ground maps with a high built density trained with sparse (left) and dense (right) street network data with a spatial resolution of 1:8000.

Varying the above mentioned factors and the resolution, we trained a model in order to generate figure-ground maps accordingly. Experimental results shown in Figure 5 highlight the differences in the generation of high density figure-ground maps using sparser or denser street network training data. Particularly, clear to see is that the model is much better to generate inner courtyards which are typical for closed building development in high built density districts in Germany when small access roads to these inner courtyards are provided to the model during training. Moreover, a denser, more detailed street network seems to obviously improve the generation of built structures that follow along the geometry of streets as highlighted by Figure 5. Interestingly, the model does not seem to get overwhelmed when learning on a denser street network as the streets are always generated superbly with only minor divergence from the input image.

Figure 6. Comparison of two generated figure-ground maps with a mid built density trained with sparse (left) and dense (right) street network data with a spatial resolution of 1:8000.

Doing the same comparison but for mid density figure-ground maps, the visual difference between the maps generated by the model that was trained on dense street network data and the one that was trained on sparse street network data is less noticeable as visualized in Figure 6.

Figure 7 example 2) demonstrates a generated figure-ground map where even individual pathways to house entrances are considered by the model and are clearly distinguishable. Furthermore, example 1) of the same figure depicts results where

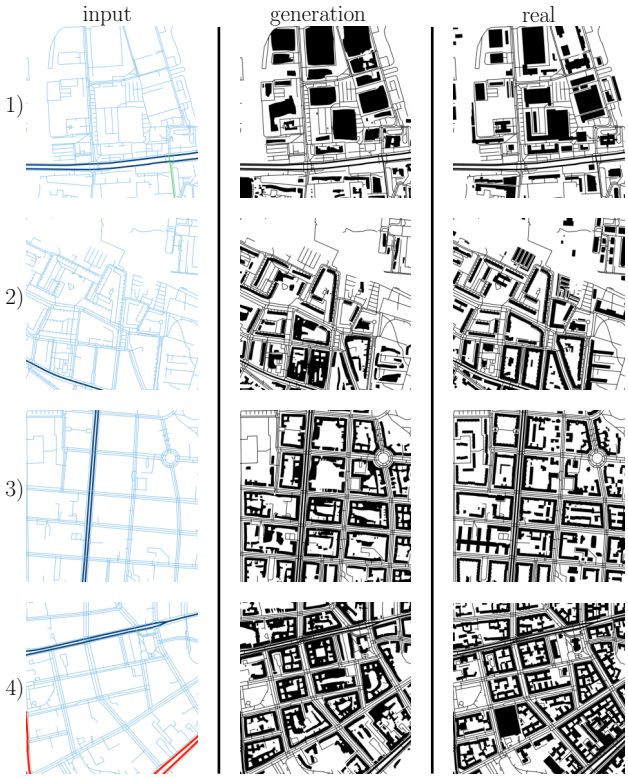

Figure 7. Results from four different high density test images with a spatial resolution of 1:8000 and a dense street network. The input image, the generated figure-ground map and the real figure-ground map are displayed.

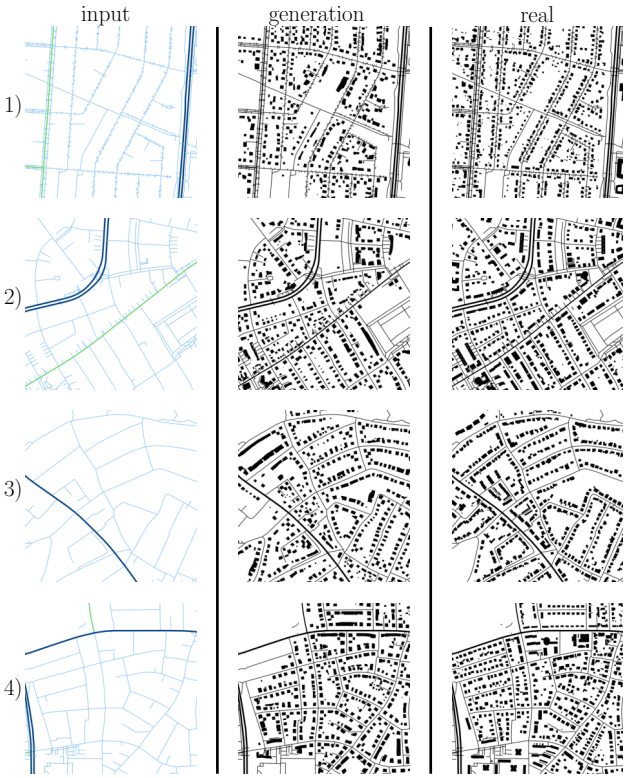

Figure 8. Results from four different mid density test images with a spatial resolution of 1:8000 and a dense street network. The input image, the generated figure-ground map and the real figure-ground map are displayed.

large industrial facilities are accurately generated due to the arrangement, type and width of the streets. For the mid built density, we can further state that the generated scenes are characterized by organic and reasonable building footprint shapes as can be depicted in Figure 8.

Analyzing the figure-ground maps for high and mid densities at a spatial resolution of 1:4000 issued from a trained model on a dense street network reveals a large discrepancy to those generated at 1:8000 as indicated by Figure 9. At a spatial resolution of 1:4000, the model seems to be incapable to produce accurate building geometries both for high and mid density scenarios.

To conclude our experimental results, we calculated the FID scores. As indicated by Table 2, figure-ground maps generated with a model that is trained on a spatial resolution of 1:4000 results in worse performance compared to the output of a model that is trained on a spatial resolution of 1:8000, supporting the findings of the visual analysis. The FID scores for the figure-ground maps resulting from high and mid densities for a spatial resolution of 1:8000 show that training the model with sparse street network data hampers the performance of the model significantly.

density	street network	spatial resolution	FID
high	dense	1:4000	153.14
mid	dense	1:4000	140.20
high	dense	1:8000	71.25
mid	dense	1:8000	75.52
high	sparse	1:8000	116.46
mid	sparse	1:8000	84.77

Table 2. FID scores for the conducted experiments.

Figure 9. Results from four different high and mid density test images with a spatial resolution of 1:4000 and dense street network data.

5. Conclusion and Outlook

This paper introduced an approach towards a deep automatic generation of figure-ground maps commonly used in urban plan-

ning. We show that our approach is able to automatically generate figure-ground maps that can be adjusted for high and mid built densities. The generated maps consist of artificially generated building footprints and a predefined street network. This opens up new opportunities to support the as yet tedious site plan drafting and has the potential to accelerate this time consuming task which different stages of urban development rely on. The trained model turns out to capture the built density which has been reflected by the resulting maps. In this context, a training and data processing on a dense and more detailed street network with a mid spatial resolution impacted the resulting arrangement and sharpness of building footprints. As of now the models backbone is an unmodified conditional GAN. A goal of future research is the incorporation of further typical site plan requirements, e.g. the physical dimensions of a plot or constraints on the ratio between built and unbuilt area. For a more accurate representation of the population and built densities, the incorporation of the desired number of floors will be addressed as well. Additionally, detailed experimental investigation including parameter tuning and optimization will be subject of future work.

Acknowledgements

The authors would like to express their gratitude to the open code from Wu and Biljecki (2023).

References

BMVI, 2018. Regionalstatistische Raumtypologie (RegioStar) des BMVI für die Mobilitäts- und Verkehrsforschung. Federal Ministry of Transport and Digital Infrastructure. bmdv.bund.de (15 February 2024).

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.

European Commission, 2016. Country summary of Germany. Testing the degree of urbanisation at the global level. ghsl.jrc.ec.europa.eu (15 February 2024).

Gong, Q., Li, J., Liu, T., Wang, N., 2020. Generating urban fabric in the orthogonal or non-orthogonal urban landscape. *Environment and Planning B: Urban Analytics and City Science*, 47(1), 25–44.

Har-Peled, S., Raichel, B., 2014. The Fréchet distance revisited and extended. *ACM Transactions on Algorithms (TALG)*, 10(1), 1–22.

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. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 6629–6640.

Hijazi, I., Li, X., Koenig, R., Schmit, G., El Meouche, R., Lv, Z., Abune'meh, M., 2017. Measuring the homogeneity of urban fabric using 2D geometry data. *Environment and Planning B: Urban Analytics and City Science*, 44(6), 1097–1121.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967–5976.

Jiang, F., Ma, J., Webster, C. J., Chiaradia, A. J., Zhou, Y., Zhao, Z., Zhang, X., 2023a. Generative urban design: A systematic review on problem formulation, design generation, and decision-making. *Progress in Planning*, 100795.

Jiang, F., Ma, J., Webster, C. J., Li, X., Gan, V. J., 2023b. Building layout generation using site-embedded GAN model. *Automation in Construction*, 151, 104888.

Jiang, F., Ma, J., Webster, C. J., Wang, W., Cheng, J. C., 2024. Automated site planning using CAIN-GAN model. *Automation in Construction*, 159, 105286.

Statista, 2024. Statista. de.statista.com (15 February 2024).

Stiny, G., 1980. Introduction to shape and shape grammars. *Environment and planning B: planning and design*, 7(3), 343–351.

Wang, J., Huang, W., Biljecki, F., 2024. Learning visual features from figure-ground maps for urban morphology discovery. *Computers, Environment and Urban Systems*, 109, 102076.

Wu, A. N., Biljecki, F., 2022. GANmapper: geographical data translation. *International Journal of Geographical Information Science*, 36(7), 1394-1422.

Wu, A. N., Biljecki, F., 2023. InstantCITY: Synthesising morphologically accurate geospatial data for urban form analysis, transfer, and quality control. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195, 90–104.