Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

A review of computer graphics approaches to urban modeling from a machine learning perspective

Tian FENG[†‡1], Feiyi FAN[2], Tomasz BEDNARZ[3,4]

[1]*Department of Computer Science and Information Technology, La Trobe University, VIC 3086, Australia*
[2]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China*
[3]*Expanded Perception and Interaction Centre, University of New South Wales, NSW 2021, Australia*
[4]*CSIRO's Data61, NSW 2015, Australia*
[†]E-mail: t.feng@latrobe.edu.au

Received Apr. 1, 2020; Revision accepted Aug. 7, 2020; Crosschecked May 7, 2021; Published online May 22, 2021

Abstract: Urban modeling facilitates the generation of virtual environments for various scenarios about cities. It requires expertise and consideration, and therefore consumes massive time and computation resources. Nevertheless, related tasks sometimes result in dissatisfaction or even failure. These challenges have received significant attention from researchers in the area of computer graphics. Meanwhile, the burgeoning development of artificial intelligence motivates people to exploit machine learning, and hence improves the conventional solutions. In this paper, we present a review of approaches to urban modeling in computer graphics using machine learning in the literature published between 2010 and 2019. This serves as an overview of the current state of research on urban modeling from a machine learning perspective.

Key words: Urban modeling; Computer graphics; Machine learning; Deep learning
https://doi.org/10.1631/FITEE.2000141 **CLC number:** TP399

1 Introduction

Since the founding of the United Nations, a significant urbanization process has reshaped the distribution of the world's population. In 2018, 55% of the people on this planet were living in urban areas, and the proportion will probably increase to 68% by the middle of the century (United Nations, 2018). The trend transformed cities into an integral part of human society, and therefore the demand for visualizing, simulating, and perceiving urban areas is growing for numerous purposes. For instance, city planners seek urban design tools to sketch plausible blueprints for new towns according to policies, government decision-makers depend on socioeconomic information to provide high-quality public services, and digital media professionals endeavor to create virtual cities to produce vivid simulations. These stakeholders continue to discover practical solutions to urban modeling, which aims to generate the physical structure and appearance of a city.

A city demonstrates its spatiotemporal alterations in socioeconomic and cultural development. The high visual and functional complexity increases the difficulty of urban modeling using non-computational approaches. In the area of computer graphics, it is conventional to model analogous complex systems using computational approaches, such as a parallel rewriting system for synthesizing detailed plant models that exchange information with the environment (Měch and Prusinkiewicz, 1996). Consequently, computer graphics researchers have exploited similar approaches for generating urban layouts (Lipp et al., 2011; Vanegas et al., 2012b; Yang et al., 2013; Garcia-Dorado et al., 2014; Peng

‡ Corresponding author
ORCID: Tian FENG, https://orcid.org/0000-0001-9691-3266

et al., 2016; Wu WM et al., 2018) and modeling landscape architectures (Vanegas et al., 2010; Musialski et al., 2012; Bao et al., 2013a; Wu FZ et al., 2014; Lienhard et al., 2017), which achieved remarkable successes. It draws our attention that they are yet not to completely outperform manual approaches regarding controllability and accuracy, due to the high complexity of cities.

Along with the burgeoning development of computational power, a series of breakthroughs in artificial intelligence (AI), especially one of its principal branches, machine learning, have enabled efficient approximation of non-linear relationships and prediction of trends. These contributions facilitate urban modeling in terms of manipulating rules and parameters to achieve satisfactory results. For example, pre-trained machine learning models contribute to discovering layout patterns in the real world and reusing them to satisfy design requirements (Merrell et al., 2010; Vanegas et al., 2012a; Feng et al., 2016). Nevertheless, we have not found a significant survey about the participation of machine learning in urban modeling.

This paper provides a literature review on computational approaches to urban modeling from a machine learning perspective. Instead of enumerating the entire collection of related works, we intentionally limited the review to research articles in the area of computer graphics, which were published between 2010 and 2019, while machine learning was acquiring widespread adaption and popularity.

2 Algorithms in machine learning

Machine learning is known for its focus on data-driven analysis and prediction by constructing computational models from user-given inputs, and has direct relationships with statistics and mathematical optimization. The rest of this section provides a brief overview of the representative algorithms in machine learning. We recommend several classics (Bishop, 2006; Hastie et al., 2009; James et al., 2014) for readers interested in a detailed introduction to machine learning.

2.1 Categorization

The major types of machine learning algorithms include supervised learning (Caruana and Niculescu-Mizil, 2006), unsupervised learning (Khanum et al., 2015), and reinforcement learning (Kaelbling et al., 1996). The first two types differ in the use of training data with pre-existing labels. Specifically, a supervised learning algorithm aims to learn function mapping input data to an output label based on input-output pairs for classification or regression, whereas an unsupervised learning algorithm models probability densities over inputs without pre-existing labels for clustering and dimensionality reduction. As a variant related to supervised and unsupervised learning, semi-supervised learning (Zhu and Goldberg, 2009) takes advantage of the partially labeled data during training. In contrast to focusing on whether the data is labeled, a reinforcement learning algorithm is concerned with the paradigm, in which agents take actions in an environment to maximize rewards.

2.2 Traditional algorithms

The following traditional algorithms have been widely used in the area of machine learning: In supervised learning, the support vector machine (SVM) algorithm (Cortes and Vapnik, 1995) produces a gap with the maximum margin of separation to distinguish training data for classifying test data, where a kernel enabled a non-linear hyperplane as the gap (Scholkopf and Smola, 2001). The k-nearest neighbor (k-NN) algorithm (Cover and Hart, 1967) predicts the label of a test example from its k closest training examples in the feature space. The random forest (RF) algorithm (Ho, 1995) builds multiple decision trees in randomly selected domains of the feature space. The adaptive boosting (AdaBoost) algorithm (Schapire, 1999) converts a set of weak learners into a strong one. In unsupervised learning, the k-means algorithm (MacQueen, 1967) presumes k clusters in the data and partitions each example into one with the nearest centroid in the feature space, and such cluster centroids are repeatedly updated until the process converges.

2.3 Deep learning algorithms

In the last decade, the astonishing breakthroughs in high-performance computing enabled general-purpose computing on graphics processing units, and consequently revived several underestimated machine learning techniques (Fukushima, 1980; Hopfield, 1982) invented in the 1980s. These techniques were derived from neural networks (NNs) (Hassoun, 1995) that follow the information processing mechanism in biological systems. Researchers have hence proposed numerous multi-layered architectures based on NNs, which were summarized as deep learning (Goodfellow et al., 2016). They differ substantially from the traditional machine learning algorithms, because features can be discovered at various levels of abstraction from data and processed consecutively by layers, instead of being task-specific. As a prevailing deep learning algorithm, convolutional neural networks (CNNs) (Lecun et al., 1998) are dominating the area of computer vision and pattern recognition (Krizhevsky et al., 2017). Among all the hidden layers in CNNs, convolutional layers use neurons to localize visual features and spatial coherences by convolution operations on the data in the receptive fields. In comparison, autoencoders (AEs) (Rumelhart et al., 1986) comprise an encoder NN for code generation and a decoder NN for input re-instantiation from the code, which enable dimensionality reduction. The more recently introduced generative adversarial networks (GANs) (Goodfellow et al., 2014) imitates a real-world data distribution and generates similar examples from random input noise. They are based on a zero-sum game and contain two adversarial neural networks: the generator neural network produces fake examples that are indistinguishable to the discriminator neural network, which meanwhile recognizes real ones in the training data. As a kind of supervised extension of GANs, conditional generative adversarial networks (Mirza and Osindero, 2014) train the generator and the discriminator with additional labels, and are specialized for image-to-image synthesis (Isola et al., 2016).

3 Problems in urban modeling

As an attractive direction in the area of computer graphics, urban modeling spans plenty of research problems. In a comprehensive review (Aliaga, 2012), research on the three-dimensional (3D) design of smart cities was divided into geometrical modeling and behavioral modeling. In contrast, our review focuses on urban structure and appearance. According to the classification of urban elements which enable a city to be visibly organized and sharply identified (Lynch, 1964), we categorize the related problems into two application classes: layout modeling and architectural modeling. In particular, the layout modeling stage generates hierarchical paths for movement and subdivides the urban area into finer parcels for placing architecture; the architectural modeling stage synthesizes geometries and textures to represent shapes and facades of buildings.

We collected in total 50 research articles related to the above-mentioned application classes, which were published between 2010 and 2019, using Google Scholar. Table 1 lists them by application class and participation of machine learning. Most of these articles were accepted by outstanding journals (e.g., *ACM Transactions on Graphics* and *Computer Graphics Forum*) and conferences (e.g.,

Table 1 Research articles in the review by application class and participation of machine learning

	Layout modeling	Architectural modeling
With machine learning	Vanegas et al. (2012a), Hartmann et al. (2017)	Lafarge and Mallet (2011), Lin et al. (2013), Demir et al. (2014), Guerrero et al. (2015), Nan et al. (2015), Affara et al. (2016), Nishida et al. (2016b), Kelly et al. (2017, 2018), Kim et al. (2020), Newton (2019)
Without machine learning	Galin et al. (2010, 2011), Lipp et al. (2011), Emilien et al. (2012), Vanegas et al. (2012b), Yu and Steed (2012), Yang et al. (2013), Beneš et al. (2014), Garcia-Dorado et al. (2014, 2017), Peng et al. (2014, 2016), Nishida et al. (2016a), Fernandes and Fernandes (2018), Mathew et al. (2019)	Krecklau et al. (2010), Nan et al. (2010), Vanegas et al. (2010), Zheng et al. (2010), Shen et al. (2011), Ceylan et al. (2012), Musialski et al. (2012), AlHalawani et al. (2013), Bao et al. (2013a, 2013b), Besuievsky and Patow (2013), Lin et al. (2013), Zhang et al. (2013), Ceylan et al. (2014), Dang et al. (2014), Wu FZ et al. (2014), Ilčík et al. (2015), Kelly et al. (2015), Schwarz and Müller (2015), Li ML et al. (2016), Lienhard et al. (2017), Smith et al. (2018)

ACM SIGGRAPH, ACM SIGGRAPH Asia and Eurographics) in the area of computer graphics. Fig. 1 indicates the number of research articles in the review by application class and year, which suggests the increasing attention to architectural modeling rather than layout modeling. We observed that machine learning was not to overshadow conventional approaches in urban modeling, but acted as an auxiliary component in limited works. However, its participation was expanding in recent years. Fig. 2 demonstrates this trend with the number of research articles in our review by participation of machine learning and year. Symbolic research articles in urban modeling, which do not involve machine learning, are also discussed in the review for comparison.

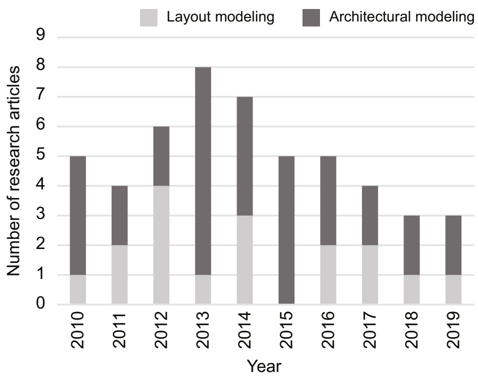

Fig. 1 Number of research articles by application class and year

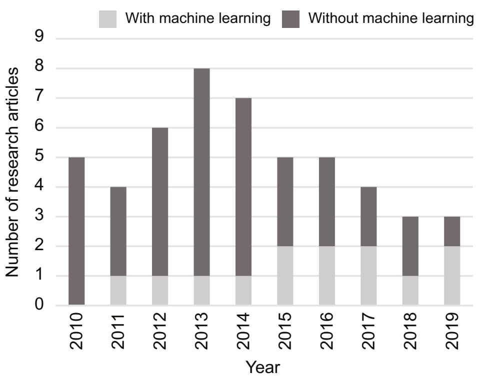

Fig. 2 Number of research articles by participation of machine learning and year

4 Layout modeling

As the two-dimensional (2D) representation of a city's physical structure, an urban layout comprises two parts: paths and parcels. The hierarchical paths constitute a transport network that serves pedestrians and vehicles with different capacities (e.g., intercity highways, intracity arteries, and residential streets). Upon the completion of the city's skeleton, the transport network divides the urban area into city blocks, which are the smallest zones surrounded by paths and form the urban fabric as basic units (Lynch, 1964). Each city block is then subdivided into a set of finer parcels. Following a land-use plan, a parcel can be owned publicly or privately for constructing only one building in most cases. In the area of computer graphics, the mainstream approaches to layout modeling are based on rules, examples, or objectives.

4.1 Rule-based approaches

Rule-based approaches to layout modeling were derived from procedural modeling, a set of techniques that allow fast production of specific contents from input parameters and generative rules. They have been exploited in the synthesis of textures and virtual worlds (Smelik et al., 2014).

Parish and Müller (2001) proposed the pioneering work with procedural modeling for urban modeling. Specifically, they used the L-system (Lindenmayer, 1968), a parallel rewriting system, for plant modeling. Thereafter, Galin et al. (2010) introduced a procedural approach to road generation, which enables the construction of bridges and tunnels; the weighted anisotropic shortest path algorithm in the approach considers impacts of water bodies, mountains, and plants. Galin et al. (2011) extended it with a subsequent attempt at a hierarchical road network that connects cities and towns. With a focus on topological validity, Lipp et al. (2011) proposed a procedural urban layout editing method using a layering system. Emilien et al. (2012) presented a progressive method of village generation; it places seeds iteratively following multiple criteria and creates roads given an input terrain, and each seed expands to a parcel. Instead of the L-system, Fernandes and Fernandes (2018) presented an alternative approach to road layout generation inspired by another plant generation algorithm, space colonization, in which attraction points were used to expedite the parameterization of a road layout.

Some researchers began to consider temporal factors in rule-based approaches via simulation. Vanegas et al. (2012b) proposed an algorithm that partitions a city block according to a simulation progress along with parameter values; it starts with the oriented bounding box of the block and subdivides the block iteratively based on the straight skeleton of a generated polygon. Likewise, Beneš et al. (2014) introduced a procedural method that grows an urban layout over time by street expansion and block subdivision with traffic and land-use simulations; it contemplates the implications of water transportation and neighboring cities. Besides generation, Garcia-Dorado et al. (2014) devised a traffic micro-simulation engine for procedural modeling of a road network that considers travel time and vehicle emissions, and Mathew et al. (2019) integrated inverse procedural modeling with population simulation to yield virtual urban environments with walkability.

Although the works above contributes to layout modeling, they are not to completely meet the critical challenge in procedural modeling (i.e., requirements for expertise and knowledge in the domain from users for rule design and parameter configuration). Machine learning has hence been used to address this challenge. In the inverse procedural method proposed by Vanegas et al. (2012a), user-specified indicators about high-level design goals drive the search for parameter values via a stochastic optimization algorithm based on Markov chain Monte Carlo (MCMC). It requires numerous state changes in an iterative procedural system and leads to huge computation against interactivity. Therefore, the authors applied a multi-layer NN trained on different ranges of parameter values and urban scenarios to approximate the original search. This method with machine learning brings a substantial improvement in efficiency.

4.2 Example-based approaches

To circumvent the challenge in rule-based approaches, researchers analyzed real-world examples for layout modeling. In particular, they focused on characterizing existent data and reusing it for robust synthesis. Yu and Steed (2012) proposed a road network synthesis method that seeks the best real-world counterpart regarding topological similarity at unfinished nodes via neighborhood edge matching; it ensures that the generated roads satisfy specific constraints, such as obstacle avoidance.

An example-based approach can collaborate with a procedural model for realistic urban details. Nishida et al. (2016a) introduced an interactive system for urban road modeling; it specifies example networks and extracts patches and statistical features to drive procedural modeling without configuring parameters.

Distinguished from the above works, a revolutionary approach to road network generation proposed by Hartmann et al. (2017) involves deep learning, GANs in particular. The authors trained GANs on a set of rasterized real-world streets to synthesize an image from input noise, and processed it into a road network via a graph-based representation. It avoids the extraction of example characteristics and compliance with constraints, but is capable of generating structurally sound and visually reasonable road networks.

4.3 Objective-based approaches

A significant concern of modeling tasks is to ensure that results satisfy user-specified high-level goals, which can be regarded as objectives. Consequently, researchers solved layout modeling via optimization according to the explicitly formulated objective functions. Such an objective function comprises multiple cost terms that address design goals so as to drive the optimization, which edits an urban layout iteratively considering specific constraints.

Yang et al. (2013) proposed a hierarchical splitting method for the generation of roads and parcels; it includes a global optimization stage based on a sparse linear system that improves the quality of layouts regarding block regularity and road fairness, in addition to a coarse-scale streamline-based splitting and a fine-scale template-based splitting. Peng et al. (2014) introduced a layout generation framework based on domain tiling with deformable templates; it focuses on accessibility and aesthetics via integer programming, a deterministic linear optimization algorithm. The authors extended it over the input's quad mesh and enabled the optimization algorithm with respect to construction and travel costs (Peng et al., 2016).

We failed to find an objective-based approach to layout modeling with machine learning. In Section 4.1, we discussed a method proposed by Vanegas et al. (2012a) using machine learning to approximate the procedural modeling parameters, which should have been obtained via optimization, instead of the output layouts. The similar strategy of employing machine learning for approximation was found in several objective-based approaches to other layout generation problems. Merrell et al. (2010) proposed an MCMC-based optimization method for residential floorplan design; it exploits Bayesian network, a supervised learning algorithm, towards semantic structures in real-world examples; the trained Bayesian network produces an architectural program, which encodes a floorplan's specification as the input to the optimization that edits the floorplan regarding shape, floor, dimension, and accessibility. Feng et al. (2016) introduced another MCMC-based optimization method for mid-scale layout design (e.g., shopping malls and train stations); it relies on RFs for the instant approximation of cost terms concerning mobility, accessibility, and coziness, which should have been computed by crowd simulation in a time-consuming manner.

5 Architectural modeling

The style of a virtual urban environment relates predominantly to the architectural models that are populated following layout modeling. An architectural model is constituted by a 3D geometry for its shape, and 2D textures for its facade (i.e., the front of a building that faces a street or open space). Architectural modeling usually works with repeating patterns observed in the real world. The mainstream approaches to architectural modeling are based on rules or reconstructions in the area of computer graphics.

5.1 Rule-based approaches

Some early works (Wonka et al., 2003; Müller et al., 2006) proved the popularity and strength of rule-based approaches in architectural modeling as that in layout modeling. Thereafter, Vanegas et al. (2010) proposed a shape grammar that progressively produces architectural models from calibrated aerial images. Krecklau et al. (2010) presented a procedural modeling language, G^2 (generalized grammar), that generates architectural models with high descriptive power based on the concepts of general-purpose programming languages. Bao et al. (2013a) integrated heuristic search and quadratic programming for grammar-based hierarchical segmentation and labeling that generate facades similar to given examples. The authors extended the approach with analysis of successful real-world architectures so that it could characterize the space of location variations and retain goodness (Bao et al., 2013b). Besuievsky and Patow (2013) devised a level-of-detail (LOD) user specification for model reduction that helps procedural modeling of architectures; it enables artists to create an architectural model without any programming rule. Wu FZ et al. (2014) proposed an inverse method for grammar extraction from a given layout via approximate dynamic programming; it is driven by rule type and symbol sequence length. Ilčík et al. (2015) introduced a facade generation method via multiple overlapping layers that describe facade elements with compact generator patterns; it solves the interactivity problem due to the complexity of facades in tree-based procedural approaches. Kelly et al. (2015) proposed a guideline-based algorithm that defines dimensioning lines for procedural models, which supports architectural modeling. Schwarz and Müller (2015) presented a grammar language, CGA++, that extended a previous work of CGA (Müller et al., 2006); it aims at procedural modeling of architecture and overcomes several limitations, such as the coordination of refinement decisions and operations involving multiple shapes. Lienhard et al. (2017) introduced a set of procedural rule-merging algorithms for fine-grained variations for architectural models.

An outstanding rule-based work with traditional unsupervised learning was proposed by Demir et al. (2014), which converts an unlabeled 3D architectural model into a procedural representation via similarity-based de-instancing and repetitive pattern discovery; it performs a dissimilarity clustering of input models using the k-means algorithm and determines the types of the hierarchical components.

Regarding the use of deep learning in rule-based approaches, Nishida et al. (2016b) devised a procedural architectural modeling method with two CNNs that finds the best pre-defined grammar, snippet, matching a user-input sketch with corresponding parameter values. More recently, Kelly et al. (2018) proposed FrankenGAN, a GAN-based interactive system that creates plausible details over procedurally generated mass models at different scales and produces consistent style distributions over buildings and neighborhoods guided by exemplar images. In an inverse procedural modeling system, Kim et al.

(2020) employed GANs to create terrain and height maps and CNNs to identify spatial properties of component arrangements; the method can create a city model that matches the style of an input street-view image.

5.2 Reconstruction-based approaches

Nowadays, the acquisition of heterogeneous urban data costs much less since data collection devices (e.g., laser scanners and RGB-D cameras) become inexpensive. A very recent example is that Apple equipped the latest model of its flagship tablet iPad Pro with a LiDAR scanner (Aleotti et al., 2020). This trend enables an alternative solution to architectural modeling via the reconstruction of shapes and facades from images and point clouds.

Nan et al. (2010) introduced an interactive tool that supports quick assembling of an architectural model over a 3D point cloud from large-scale scanning of an urban scene. Zheng et al. (2010) proposed an urban scene reconstruction method with non-local filtering by scanning LiDAR data repeatedly and completing the missing parts. Lafarge and Mallet (2011) solved semantic architectural reconstruction as non-convex energy minimization with vegetation and complex grounds from unstructured points. Shen et al. (2011) presented an adaptive facade partitioning method over LiDAR data; it outputs sub-facades that are described by various rectilinear grids of patterns and can be merged following their similarities for further operations. Ceylan et al. (2012) presented an image-based framework for architecture reconstruction; it uses the symmetry priors from 3D input lines for edge detection. The authors further proposed a regularity detection method for calibrated urban facades using graph-based optimization, which can also create 3D scenes (Ceylan et al., 2014). AlHalawani et al. (2013) introduced a method that extracts the representation of factored facades based on repetitive patterns and deformations at the level of windows; the output encodings facilitate the interactive facade generation. Kuang et al. (2013) proposed an efficient structure to render detailed architectural models via non-uniform grid-based subdivision of images and point clouds. Zhang et al. (2013) devised a layered generative method that analyzes structures of irregular facades and simplifies their representations. Dang et al. (2015) presented an image-based facade-editing framework that exploits topological jump and spatial optimization for discrete and continuous modifications. Li ML et al. (2016) partitioned the point cloud into a non-uniform grid to obtain well-aligned boxes and then approximate an architectural geometry; based on a novel Markov random field (MRF) formulation, it supports multiple types of point clouds. A recent work proposed by Smith et al. (2018) targets urban reconstruction from multi-view stereo and addresses the challenge of automatic view and path planning for aerial imaging based on unmanned aerial vehicles; it is based on continuous optimization using heuristics for multi-view stereo reconstruction and enables quick generation of urban paths for large scan areas.

The relationship between reconstruction and computer vision tasks (i.e., semantic segmentation and object recognition) enables the extensive participation of machine learning. In the reconstruction-based approaches to architectural modeling, unsupervised learning algorithms contribute to preprocessing of unlabeled raw data and diminishing the difficulty of reconstruction, which is similar to a rule-based approach proposed by Demir et al. (2014). Musialski et al. (2012) introduced a coherence-based facade-generation method that uses a hierarchical agglomerative clustering algorithm for image splitting. Verdie et al. (2015) proposed an LOD method that creates semantic architectural geometries and performs an MRF clustering of super-facets from the input mesh prior to reconstruction.

In comparison, supervised learning algorithms are preferred upon the availability of labeled data. Xiao et al. (2009) and Lin et al. (2013) exploited the AdaBoost algorithm to classify image pixels and LiDAR data points for reconstructing street-side and residential scenes. Guerrero et al. (2015) presented a facade-generation method that learns and propagates shape placements from user-given examples using kernel regression. Nan et al. (2015) employed linear regression to predict weights in template assembly optimization, which reconstructs the shape details of architectural models. Affara et al. (2016) introduced SVMs trained on an urban image dataset for the iterative update of priors, which guide the detection of facade elements; the same problem was solved by Kelly et al. (2017) via CNNs and the fusion of heterogeneous data. In a recent survey, Newton (2019) analyzed a variety of GAN-based methods about their contributions to the creation and design of architectural models from specific styles.

6 Limitations and opportunities

Among the research articles in this review, we observed that several of which resorted to unsupervised learning algorithms instead of the preferred supervised learning algorithms. It is likely due to the shortage of meaningfully labeled urban data. Although manual labeling consumes labor and causes this situation, it could be assisted by task-specific NNs, such as places-CNN (Zhou B et al., 2014), for urban scene recognition. Hidden layers in such pretrained models can serve as labels for supervision and features for clustering. Alternatively, people may obtain labels from geo-tagged census data using geo-spatial methods (Hu et al., 2015).

An urban modeling task is usually associated with a huge solution space, especially when its solution depends on rules and objectives with non-linear relationships and high-dimensional parameters. The resulting complexity would increase the cost of computation. Machine learning is believed to address such a gap in various ways. For example, the design space of a procedural model can be downsized to a lower-dimensional one using AEs (Yumer et al., 2015); CNNs enable the efficient estimation of parameter values (Huang et al., 2017); the search space of optimization shrinks if prior information (Merrell et al., 2010) and initial arrangement (Vanegas et al., 2012a) are provided via hierarchical machine learning algorithms.

Some early works attempted to solve urban modeling problems with pre-defined behavioral data (Vanegas et al., 2009) and high-level indicators (Vanegas et al., 2012a). However, they were oversimplified compared to the real-world complexity. In the area of computer vision, we noticed that research on urban perception (i.e., understanding the factors, patterns, and mechanisms of urban areas) has increased as well in recent years (Guo et al., 2017; Feng et al., 2018; Goldblatt et al., 2018; Li HN et al., 2018). This topic was expected to provide sophisticated and in-depth information for urban modeling. Although it is still hard to find articles that seamlessly integrate both topics, we believe that such exploration is worth further research attention.

7 Conclusions

The complexity of cities and the increasing demand for urban-related purposes motivate research in urban modeling regarding the effectiveness and cost-efficiency. To face several challenges, researchers have attempted different types of approaches to urban modeling, and discovered that, for the implication of solution spaces, the use of machine learning could facilitate problem solving. We reviewed 50 related research articles published between 2010 and 2019 in the area of computer graphics with a focus on the participation of machine learning. The approaches proposed in these articles were briefly introduced and categorized by application class. In addition, we discussed the observed limitations, challenges, and corresponding opportunities in urban modeling from a machine learning perspective.

Contributors

Tian FENG designed the review and collected the literature. Tian FENG and Feiyi FAN drafted the manuscript. Tian FENG, Feiyi FAN, and Tomasz BEDNARZ revised and finalized the paper.

Compliance with ethics guidelines

Tian FENG, Feiyi FAN, and Tomasz BEDNARZ declare that they have no conflict of interest.

References

Affara L, Nan LL, Ghanem B, et al., 2016. Large scale asset extraction for urban images. Proc 14^{th} European Conf on Computer Vision, p.437-452. https://doi.org/10.1007/978-3-319-46487-9_27

Aleotti F, Zaccaroni G, Bartolomei L, et al., 2020. Real-time single image depth perception in the wild with handheld devices. https://arxiv.org/abs/2006.05724

AlHalawani S, Yang YL, Liu H, et al., 2013. Interactive facades analysis and synthesis of semi-regular facades. *Comput Graph Forum*, 32(2pt2):215-224. https://doi.org/10.1111/cgf.12041

Aliaga DG, 2012. 3D design and modeling of smart cities from a computer graphics perspective. *Int Schol Res Notic*, 2012:728913. https://doi.org/10.5402/2012/728913

Bao F, Schwarz M, Wonka P, 2013a. Procedural facade variations from a single layout. *ACM Trans Graph*, 32(1):8:1-8:13. https://doi.org/10.1145/2421636.2421644

Bao F, Yan DM, Mitra NJ, et al., 2013b. Generating and exploring good building layouts. *ACM Trans Graph*, 32(4):122:1-122:10. https://doi.org/10.1145/2461912.2461977

Beneš J, Wilkie A, Křivánek J, 2014. Procedural modelling of urban road networks. *Comput Graph Forum*, 33(6):132-142. https://doi.org/10.1111/cgf.12283

Besuievsky G, Patow G, 2013. Customizable LOD for procedural architecture. *Comput Graph Forum*, 32(8):26-34. https://doi.org/10.1111/cgf.12141

Bishop CM, 2006. Pattern Recognition and Machine Learning. Springer, New York, USA.

Caruana R, Niculescu-Mizil A, 2006. An empirical comparison of supervised learning algorithms. Proc 23rd Int Conf on Machine Learning, p.161-168. https://doi.org/10.1145/1143844.1143865

Ceylan D, Mitra NJ, Li H, et al., 2012. Factored facade acquisition using symmetric line arrangements. *Comput Graph Forum*, 31(2pt3):671-680. https://doi.org/10.1111/j.1467-8659.2012.03046.x

Ceylan D, Mitra NJ, Zheng Y, et al., 2014. Coupled structure-from-motion and 3D symmetry detection for urban facades. *ACM Trans Graph*, 33(1):2. https://doi.org/10.1145/2517348

Cortes C, Vapnik V, 1995. Support-vector networks. *Mach Learn*, 20(3):273-297. https://doi.org/10.1007/BF00994018

Cover T, Hart P, 1967. Nearest neighbor pattern classification. *IEEE Trans Inform Theory*, 13(1):21-27. https://doi.org/10.1109/TIT.1967.1053964

Dang M, Ceylan D, Neubert B, et al., 2014. Safe: structure-aware facade editing. *Comput Graph Forum*, 33(2):83-93. https://doi.org/10.1111/cgf.12313

Dang M, Lienhard S, Ceylan D, et al., 2015. Interactive design of probability density functions for shape grammars. *ACM Trans Graph*, 34(6):206. https://doi.org/10.1145/2816795.2818069

Demir I, Aliaga DG, Benes B, 2014. Proceduralization of buildings at city scale. Proc 2nd Int Conf on 3D Vision, p.456-463. https://doi.org/10.1109/3DV.2014.31

Emilien A, Bernhardt A, Peytavie A, et al., 2012. Procedural generation of villages on arbitrary terrains. *Vis Comput*, 28(6-8):809-818. https://doi.org/10.1007/s00371-012-0699-7

Feng T, Yu LF, Yeung SK, et al., 2016. Crowd-driven mid-scale layout design. *ACM Trans Graph*, 35(4):132. https://doi.org/10.1145/2897824.2925894

Feng T, Truong QT, Nguyen DT, et al., 2018. Urban zoning using higher-order Markov random fields on multi-view imagery data. Proc 15th European Conf on Computer Vision, p.627-644. https://doi.org/10.1007/978-3-030-01237-3_38

Fernandes GD, Fernandes AR, 2018. Space colonisation for procedural road generation. Proc Int Conf on Graphics and Interaction, p.1-8. https://doi.org/10.1109/ITCGI.2018.8602928

Fukushima K, 1980. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*, 36(4):193-202. https://doi.org/10.1007/BF00344251

Galin E, Peytavie A, Maréchal N, et al., 2010. Procedural generation of roads. *Comput Graph Forum*, 29(2):429-438. https://doi.org/10.1111/j.1467-8659.2009.01612.x

Galin E, Peytavie A, Guérin E, et al., 2011. Authoring hierarchical road networks. *Comput Graph Forum*, 30(7):2021-2030. https://doi.org/10.1111/j.1467-8659.2011.02055.x

Garcia-Dorado I, Aliaga D G, Ukkusuri S V, 2014. Designing large-scale interactive traffic animations for urban modeling. *Comput Graph Forum*, 33(2):411-420. https://doi.org/10.1111/cgf.12329

Garcia-Dorado I, Aliaga DG, Bhalachandran S, et al., 2017. Fast weather simulation for inverse procedural design of 3D urban models. *ACM Trans Graph*, 36(2):21. https://doi.org/10.1145/2999534

Goldblatt R, Stuhlmacher MF, Tellman B, et al., 2018. Using landsat and nighttime lights for supervised pixel-based image classification of urban land cover. *Remot Sens Environ*, 205:253-275. https://doi.org/10.1016/j.rse.2017.11.026

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

Goodfellow I, Pouget-Abadie J, Mirza M, et al., 2014. Generative adversarial nets. Proc 27th Int Conf on Neural Information Processing Systems, p.2672-2680.

Guerrero P, Jeschke S, Wimmer M, et al., 2015. Learning shape placements by example. *ACM Trans Graph*, 34(4):108. https://doi.org/10.1145/2766933

Guo YJ, Chen EX, Guo Y, et al., 2017. Deep highway unit network for land cover type classification with GF-3 SAR imagery. Proc SAR in Big Data Era: Models, Methods and Applications, p.1-6. https://doi.org/10.1109/BIGSARDATA.2017.8124926

Hartmann S, Weinmann M, Wessel R, et al., 2017. Street-GAN: towards road network synthesis with generative adversarial networks. Proc 25th Int Conf in Central Europe on Computer Graphics, Visualization and Computer Vision, p.133-142.

Hassoun MH, 1995. Fundamentals of Artificial Neural Networks. MIT Press, Cambridge, USA.

Hastie T, Tibshirani R, Friedman J, 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, USA.

Ho TK, 1995. Random decision forests. Proc 3rd Int Conf on Document Analysis and Recognition, p.278. https://doi.org/10.1109/ICDAR.1995.598994

Hopfield JJ, 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA*, 79(8):2554-2558. https://doi.org/10.1073/pnas.79.8.2554

Hu YJ, Gao S, Janowicz K, et al., 2015. Extracting and understanding urban areas of interest using geotagged photos. *Comput Environ Urban Syst*, 54:240-254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Huang HB, Kalogerakis E, Yumer E, et al., 2017. Shape synthesis from sketches via procedural models and convolutional networks. *IEEE Trans Vis Comput Graph*, 23(8):2003-2013. https://doi.org/10.1109/TVCG.2016.2597830

Ilčík M, Musialski P, Auzinger T, et al., 2015. Layer-based procedural design of façades. *Comput Graph Forum*, 34(2):205-216. https://doi.org/10.1111/cgf.12553

Isola P, Zhu J, Zhou TH, et al., 2016. Image-to-image translation with conditional adversarial networks. https://arxiv.org/abs/1611.07004

James G, Witten D, Hastie T, et al., 2014. An Introduction to Statistical Learning: with Applications in R. Springer, New York, USA.

Kaelbling LP, Littman ML, Moore AW, 1996. Reinforcement learning: a survey. *J Artif Intell Res*, 4(1):237-285.

Kelly T, Wonka P, Mueller P, 2015. Interactive dimensioning of parametric models. *Comput Graph Forum*, 34(2):117-129. https://doi.org/10.1111/cgf.12546

Kelly T, Femiani J, Wonka P, et al., 2017. BigSUR: large-scale structured urban reconstruction. *ACM Trans Graph*, 36(6):204. https://doi.org/10.1145/3130800.3130823

Kelly T, Guerrero P, Steed A, et al., 2018. FrankenGAN: guided detail synthesis for building mass models using style-synchonized GANs. *ACM Trans Graph*, 37(6):216. https://doi.org/10.1145/3272127.3275065

Khanum M, Mahboob T, Imtiaz W, et al., 2015. A survey on unsupervised machine learning algorithms for automation, classification and maintenance. *Int J Comput Appl*, 119(13):34-39. https://doi.org/10.5120/21131-4058

Kim S, Kim D, Choi S, 2020. CityCraft: 3D virtual city creation from a single image. *Vis Comput*, 36(5):911-924. https://doi.org/10.1007/s00371-019-01701-x

Krecklau L, Pavic D, Kobbelt L, 2010. Generalized use of non-terminal symbols for procedural modeling. *Comput Graph Forum*, 29(8):2291-2303. https://doi.org/10.1111/j.1467-8659.2010.01714.x

Krizhevsky A, Sutskever I, Hinton GE, 2017. ImageNet classification with deep convolutional neural networks. *Commun ACM*, 60(6):84-90. https://doi.org/10.1145/3065386

Kuang ZZ, Chan B, Yu YZ, et al., 2013. A compact random-access representation for urban modeling and rendering. *ACM Trans Graph*, 32(6):172. https://doi.org/10.1145/2508363.2508424

Lafarge F, Mallet C, 2011. Building large urban environments from unstructured point data. Proc Int Conf on Computer Vision, p.1068-1075. https://doi.org/10.1109/ICCV.2011.6126353

Lecun Y, Bottou L, Bengio Y, et al., 1998. Gradient-based learning applied to document recognition. *Proc IEEE*, 86(11):2278-2324. https://doi.org/10.1109/5.726791

Li HN, Jia YH, Zhou Y, 2018. Urban expansion pattern analysis and planning implementation evaluation based on using fully convolution neural network to extract land range. *NeuroQuantology*, 16(5):814-822. https://doi.org/10.14704/nq.2018.16.5.1390

Li ML, Wonka P, Nan LL, 2016. Manhattan-world urban reconstruction from point clouds. Proc 14th European Conf on Computer Vision, p.54-69. https://doi.org/10.1007/978-3-319-46493-0_4

Lienhard S, Lau C, Müller P, et al., 2017. Design transformations for rule-based procedural modeling. *Comput Graph Forum*, 36(2):39-48. https://doi.org/10.1111/cgf.13105

Lin H, Gao JZ, Zhou Y, et al., 2013. Semantic decomposition and reconstruction of residential scenes from LiDAR data. *ACM Trans Graph*, 32(4):66. https://doi.org/10.1145/2461912.2461969

Lindenmayer A, 1968. Mathematical models for cellular interactions in development I. Filaments with one-sided inputs. *J Theor Biol*, 18(3):280-299. https://doi.org/10.1016/0022-5193(68)90079-9

Lipp M, Scherzer D, Wonka P, et al., 2011. Interactive modeling of city layouts using layers of procedural content. *Comput Graph Forum*, 30(2):345-354. https://doi.org/10.1111/j.1467-8659.2011.01865.x

Lynch K, 1964. The Image of the City. MIT Press, Cambridge, UK.

MacQueen J, 1967. Some methods for classification and analysis of multivariate observations. Proc 5th Berkeley Symp on Mathematical Statistics and Probability, p.281-297.

Mathew CDT, Knob PR, Musse SR, et al., 2019. Urban walkability design using virtual population simulation. *Comput Graph Forum*, 38(1):455-469. https://doi.org/10.1111/cgf.13585

Měch R, Prusinkiewicz P, 1996. Visual models of plants interacting with their environment. Proc 23rd Annual Conf on Computer Graphics and Interactive Techniques, p.397-410. https://doi.org/10.1145/237170.237279

Merrell P, Schkufza E, Koltun V, 2010. Computer-generated residential building layouts. *ACM Trans Graph*, 29(6):181. https://doi.org/10.1145/1882261.1866203

Mirza M, Osindero S, 2014. Conditional generative adversarial nets. https://arxiv.org/abs/1411.1784

Müller P, Wonka P, Haegler S, et al., 2006. Procedural modeling of buildings. *ACM Trans Graph*, 25(3):614-623. https://doi.org/10.1145/1141911.1141931

Musialski P, Wimmer M, Wonka P, 2012. Interactive coherence-based façades modeling. *Comput Graph Forum*, 31(2pt3):661-670. https://doi.org/10.1111/j.1467-8659.2012.03045.x

Nan LL, Sharf A, Zhang H, et al., 2010. SmartBoxes for interactive urban reconstruction. *ACM Trans Graph*, 29(4):93. https://doi.org/10.1145/1778765.1778830

Nan LL, Jiang CG, Ghanem B, et al., 2015. Template assembly for detailed urban reconstruction. *Comput Graph Forum*, 34(2):217-228. https://doi.org/10.1111/cgf.12554

Newton D, 2019. Generative deep learning in architectural design. *Technol Arch Des*, 3(2):176-189. https://doi.org/10.1080/24751448.2019.1640536

Nishida G, Garcia-Dorado I, Aliaga DG, 2016a. Example-driven procedural urban roads. *Comput Graph Forum*, 35(6):5-17. https://doi.org/10.1111/cgf.12728

Nishida G, Garcia-Dorado I, Aliaga DG, et al., 2016b. Interactive sketching of urban procedural models. *ACM Trans Graph*, 35(4):130. https://doi.org/10.1145/2897824.2925951

Parish YIH, Müller P, 2001. Procedural modeling of cities. Proc 28th Annual Conf on Computer Graphics and Interactive Techniques, p.301-308. https://doi.org/10.1145/383259.383292

Peng CH, Yang YL, Wonka P, 2014. Computing layouts with deformable templates. *ACM Trans Graph*, 33(4):99. https://doi.org/10.1145/2601097.2601164

Peng CH, Yang YL, Bao F, et al., 2016. Computational network design from functional specifications. *ACM Trans Graph*, 35(4):131. https://doi.org/10.1145/2897824.2925935

Rumelhart DE, Hinton GE, Williams RJ, 1986. Learning Internal Representations by Error Propagation. In:

Rumelhart DE, McClelland JL (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations. MIT Press, Cambridge, MA, USA.

Schapire RE, 1999. A brief introduction to boosting. Proc 16th Int Joint Conf on Artificial Intelligenc, p.1401-1406.

Scholkopf B, Smola AJ, 2001. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge, MA, USA.

Schwarz M, Müller P, 2015. Advanced procedural modeling of architecture. *ACM Trans Graph*, 34(4):107. https://doi.org/10.1145/2766956

Shen CH, Huang SS, Fu H, et al., 2011. Adaptive partitioning of urban facades. *ACM Trans Graph*, 30(6):184. https://doi.org/10.1145/2070781.2024218

Smelik RM, Tutenel T, Bidarra R, et al., 2014. A survey on procedural modelling for virtual worlds. *Comput Graph Forum*, 33(6):31-50. https://doi.org/10.1111/cgf.12276

Smith N, Moehrle N, Goesele M, et al., 2018. Aerial path planning for urban scene reconstruction: a continuous optimization method and benchmark. *ACM Trans Graph*, 37(6):183. https://doi.org/10.1145/3272127.3275010

United Nations, 2018. 2018 Revision of World Urbanization Prospects. https://un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html [Accessed on May 16, 2018]

Vanegas CA, Aliaga DG, Benes B, et al., 2009. Interactive design of urban spaces using geometrical and behavioral modeling. *ACM Trans Graph*, 28(5):111. https://doi.org/10.1145/1618452.1618457

Vanegas CA, Aliaga DG, Beneš B, 2010. Building reconstruction using Manhattan-world grammars. Proc IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.358-365. https://doi.org/10.1109/CVPR.2010.5540190

Vanegas CA, Garcia-Dorado I, Aliaga DG, et al., 2012a. Inverse design of urban procedural models. *ACM Trans Graph*, 31(6):168. https://doi.org/10.1145/2366145.2366187

Vanegas CA, Kelly T, Weber B, et al., 2012b. Procedural generation of parcels in urban modeling. *Comput Graph Forum*, 31(2pt3):681-690. https://doi.org/10.1111/j.1467-8659.2012.03047.x

Verdie Y, Lafarge F, Alliez P, 2015. LOD generation for urban scenes. *ACM Trans Graph*, 34(3):30. https://doi.org/10.1145/2732527

Wonka P, Wimmer M, Sillion F, et al., 2003. Instant Architecture. ACM Press, New York, NY, USA.

Wu FZ, Yan DM, Dong WM, et al., 2014. Inverse procedural modeling of facade layouts. *ACM Trans Graph*, 33(4):121. https://doi.org/10.1145/2601097.2601162

Wu WM, Fan LB, Liu LG, et al., 2018. MIQP-based layout design for building interiors. *Comput Graph Forum*, 37(2):511-521. https://doi.org/10.1111/cgf.13380

Xiao JX, Fang T, Zhao P, et al., 2009. Image-based street-side city modeling. *ACM Trans Graph*, 28(5):114. https://doi.org/10.1145/1618452.1618460

Yang YL, Wang J, Vouga E, et al., 2013. Urban pattern: layout design by hierarchical domain splitting. *ACM Trans Graph*, 32(6):181. https://doi.org/10.1145/2508363.2508405

Yu QZ, Steed A, 2012. Example-based Road Network Synthesis. Eurographics 2012-Short Papers. https://doi.org/10.2312/conf/EG2012/short/053-056

Yumer ME, Asente P, Mech R, et al., 2015. Procedural modeling using autoencoder networks. Proc 28th Annual ACM Symp on User Interface Software and Technology, p.109-118. https://doi.org/10.1145/2807442.2807448

Zhang H, Xu K, Jiang W, et al., 2013. Layered analysis of irregular facades via symmetry maximization. *ACM Trans Graph*, 32(4):121. https://doi.org/10.1145/2461912.2461923

Zheng Q, Sharf A, Wan GW, et al., 2010. Non-local scan consolidation for 3D urban scenes. *ACM Trans Graph*, 29(4):94. https://doi.org/10.1145/1778765.1778831

Zhou B, Lapedriza A, Xiao J, et al., 2014. Learning deep features for scene recognition using places database. Proc 27th Int Conf on Neural Information Processing Systems, p.487-495.

Zhu X, Goldberg AB, 2009. Introduction to semi-supervised learning. *Synth Lect Artif Intell Mach Learn*, 3(1):1-130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006