Manuscript

B Urban Analytics and City Science

EPB: Urban Analytics and City Science
2023, Vol. 0(0) 1–19

DOI: 10.1177/23998083231224505
journals.sagepub.com/home/epb
Sage

An inductive method for classifying building form in a city with implications for orientation

Jinmo Rhee
University of Calgary, Canada

Ramesh Krishnamurti
Carnegie Mellon University, USA

Abstract

The utilization of deep learning for form analysis facilitates the classification of an extensive number of forms based on their morphological features. A critical consideration for implementing such analysis methods in architectural or urban forms is whether building orientation should be embedded within the data. Orientation functions as a form variable significantly influenced by environmental, social, and cultural contexts within a city. In contrast to other domains where forms are extrapolated in relation to their context, in the city, domain orientation uniquely characterizes building form. In this paper, we introduce a pipeline for constructing an extensive building form dataset and scrutinizing the morphological identity of building forms, with a particular focus on the implications of building orientation as a manifestation of urban locality. Through a case study situated in Montreal, we engage in a comparative analysis employing two distinct datasets—those with orientation-embedded forms and those with orientation-normalized forms. Our research aims to investigate the typo-morphological characteristics of the building forms of the city and to examine how building orientation contributes to the identification of these traits and mirrors urban locality.

Keywords

Building form, deep learning, form classification, orientation, form data

Introduction

There is a growing interest in utilizing inductive methods to scrutinize extensive datasets of architectural and urban forms for morphological analysis (Li and Zhu 2020; Meeran and Joyce 2020; Miguel et al., 2019; Quan 2022; Rhee et al., 2019; Yao 2021). Within the spectrum of such analyses,

Corresponding author:
Jinmo Rhee, School of Architecture, Planning, and Landscape, University of Calgary, 2500 University Dr NW, PF 4181, Calgary, AB T2N 1N4, Canada.
Email: jinmo.rhee@ucalgary.ca

Data Availability Statement included at the end of the article

orientation emerges as a pivotal variable. In numerous fields where machine learning and deep learning are vigorously applied, such as medical science and autonomous driving, there prevails a tendency to either eliminate or normalize the orientation within data, thereby constructing datasets in a uniform format and obtaining standardized information. However, in stark contrast to these domains, the realm of architecture and urban studies considers orientation not merely as a residual element but as an intrinsic feature closely intertwined with regional and cultural attributes. This has historically necessitated its treatment as an embedded feature within the architecture and urban form analysis, rather than something to be excluded.

Yet, the exploration of how orientation operates within inductive models, particularly those categorizing a plethora of form data through deep learning, remains uncharted. This gap signifies an unmet need to establish definitive criteria delineating when to preserve orientation within datasets and when it should be omitted. The absence of such guidelines underscores a critical ambiguity in the methodological foundations of employing machine or deep learning techniques in this context, thereby necessitating a more thorough investigation into the roles and implications of orientation in the analysis of architectural and urban forms.

For this reason, this paper investigates the influence of building form orientation as a feature representing urban locality within the extensive analysis of building forms using deep learning. To this end, we introduce a pipeline comprising a sequence of computational methodologies designed to construct and scrutinize large datasets of building forms across a city and identify typological patterns. This process involves the application of two distinct datasets to the pipeline: one preserving the orientation of the building forms and another with normalized orientations. By extracting morphological types of building forms from each dataset, we then engage in a comparative analysis of these divergent types.

Through this analysis, we examine how the preservation or normalization of orientation has impacted data curation and processing, subsequently influencing the derivation of typologies. Based on these examinations, this research discusses the significance of orientation as a feature of urban locality. Furthermore, it speculatively addresses scenarios in which the orientation features are critical to the analysis, as well as instances where they might be feasibly excluded. This exploration not only underscores the nuanced role of orientation in the morphological analysis of urban and architectural forms but also prompts a re-evaluation of methodological approaches in incorporating or excluding such features in analytical models.

Background

Efforts to capture morphological features and distinguish types of building forms have intercrossed with computational analysis in urban morphology. Computational models have been employed to investigate the typological rules and characteristics of building envelopes (Steadman et al., 2000), to distinguish buildings shapes by different levels of representation (Ostwald and Vaughan, 2012), and to classify buildings by shapes of footprints and roofs (Schirmer and Kay 2015).

Those morphological analyses are based on measurements of size—dimensions, masses, angles, ratios, areas, and repetitiveness. These measurements are generally used to represent certain characteristics of forms, for example, compactness or light quality. In morphometry—a branch of morphology for the quantitative study of form, its variation, and patterns with other variables or factors through observation, measurement, and analysis (Bookstein 1992)—such size-based morphological analyses are collectively regarded as ‘traditional’ approaches (Webster and Sheets 2010). Traditional morphometry uses the size-based data and statistical techniques to capture morphological traits of biological subjects, classify them, and trace their changes and variations (Marcus 1990).

More recently, other diverse fields have developed methodological frameworks for analyzing 3D form data with deep learning techniques. They have enabled the uncovering of complex patterns of form beyond measurement. One of the more active fields using 3D deep learning in form study is neuroimaging in medical science. In neuroimaging, deep learning models are employed to process 3D brain form datasets from MRI (Magnetic Resonance Image). The models reveal relationships between brain form changes and aging (Dinsdale et al., 2021), accelerated measuring of brain parts (Rebsamen et al., 2020), and improved diagnosis of Alzheimer's disease (Zhang et al., 2019).

Deep learning is also used quantitatively in cosmology to analyze the unknown form of energy (Escamilla-Rivera et al. 2020), in cell biology to examine changes in size and shape of nuclear structures (Kalinin et al., 2018), and in geology to identify patterns of land use (Idgunji et al., 2021).

A common methodological approach can be observed from these examples: constructing a form dataset, training a deep neural network to abstract morphological features, and examining the features through patterning, clustering, and categorizing techniques. Similar approaches have been attempted in the use of deep learning models for architectural and urban form analysis. For example, urban structures are analyzed by clustering patterns of the street networks (Moosavi 2017), building types are investigated with a schematized form of a building through line composition (Miguel et al., 2019), and urban fabrics are re-clustered according to the types of building footprint configurations (Cai et al., 2021). But, in architecture and urban form studies, there are two bottlenecks to using deep learning: i) the weak use of 3D data, and ii) scarcely explored analytical potential.

Compared to other disciplines, for instance, neuroimaging, there have been few attempts made to analyze architectural and urban forms using 3D data formats (Newton 2019; Ren and Zheng 2020). Although these examples mentioned above demonstrate the potential for using 3D deep learning to generate new forms by capturing morphological features from large collections of 3D data, they are predominantly centered on generative studies. Their primary objective has been the development of deep learning models trained to synthesize specific shapes, leaving the analytical potential for architectural and urban forms relatively uncharted. Consequently, such research tends to treat forms as isolated data, focusing exclusively on the generation of shapes. However, in contrast to other domains, one of the most significant attributes of architectural data is its "contextness," encompassing locality.

Therefore, in this paper, we analyze how variations in data composition, influenced by building orientation—one of the quintessential features representing "contextness"—operate within deep learning models classifying voluminous form data. Furthermore, we discuss the significance of orientation within the analytical potential of such models.

Methods

A computational pipeline

From the common methodological approach observed in other disciplines, 3D deep learning can be embraced to explore urban morphology and architectural typology. Recognizing this potential, we developed a new pipeline for morphological analysis using a deep neural network to identify types of 3D building forms. The pipeline includes the establishment of a building form dataset, selection of a deep neural network model for capturing morphological features, training of the model with the dataset, distribution of the data based on their morphological features, investigation of the morphological types through clustering, and analysis of the characteristics of the identified types.

In the next section, we present an experimental case study conducted in Montreal to illustrate the application of the pipeline for building form analysis. The first step involves data processing, converting a large collection of building form data into a trainable dataset for a deep learning model. The next step is training, to develop a model to capture the morphological features of individual

buildings. The final step is analysis, where, by using clustering algorithms building forms can be classified according to these features to reveal the typological traits of the building forms. Moreover, by interpreting these traits at a city scale, it becomes possible to identify the morphological characteristics of the city space. See Figure 1.

Dataset

To uncover morphological patterns of building forms in a city, the first step in the pipeline is to construct two 3D building form datasets: orientation-embedded and -normalized. For this, we focus primarily on coverage, size, accessibility, and format of the data rather than on where buildings are located or to which city they belong.

Generally, a deep neural network is data-hungry, requiring a large amount of data to create a prediction model with good accuracy. This dependence on large amounts of data is even stronger in the case of 3D data. However, there is no rigid threshold on the amount of data that is required for the different types of problems, namely, object detection, classification, segmentation, and generation.

In order to set a threshold for the problem of capturing morphological features in analysis and generation, we examine similar types of problems in 3D deep learning, for example, in previous studies on 3D object representation, data samples ranging from 16,384 to 16,913 were used to measure the performance of deep neural networks to capture morphological features (Park et al., 2019; Wu et al., 2016). The category of problems we consider has comparable complexity, and therefore, we set the minimum data size to be 20,000 in order to capture the morphological features of building forms.

Another important criterion is compatibility of the data format. The raw data must be provided in a 3D format that is widely used with high editability making it suitable for further processing in a deep neural network. We specifically looked for files in OBJ, STL, SHP, DXF, or 3DM formats. Additionally, 3D objects in the dataset had to be watertight or closed to ensure stability for further processing.

Lastly, we take into account the level of detail (LOD) of the building forms in the 3D dataset. There are typically three or more levels of detail. 3D building forms in Open Street Map (OSM) are

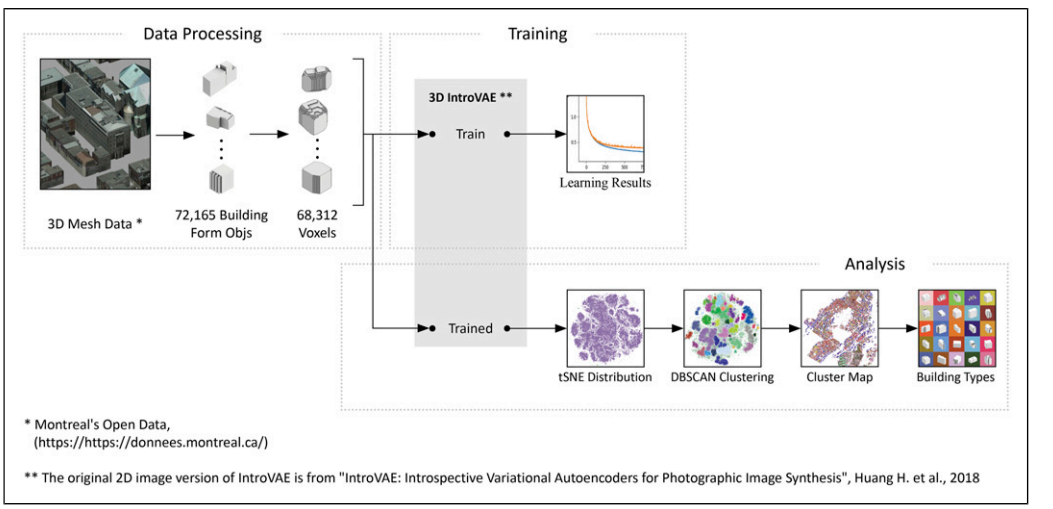

Figure 1. A pipeline for morphological analysis with a deep neural network to identify types of 3D building forms.

widely used for urban morphological research (Boeing 2017; Fleischmann et al. 2021) and meet the criteria of having a sufficient number of building forms, and of being in a compatible file format. However, 3D data in OSM is mostly simple extrusions of building footprints, which are referred to as LOD1. As a result, they lack detailed information about building features such as roof style, chimneys, podiums, and so on—these are LOD2. LOD3 represents a higher level of architectural detail than LOD2, incorporating not only building features but also openings such as windows and doors (Biljecki et al. 2016).

In this paper, we restrict the analysis to LOD2 forms. To obtain a large collection of LOD2 3D building form data, we rely on public datasets from municipal governments in CityGML, an open platform for storing and exchanging virtual 3D city models. There are a variety of building form datasets in CityGML to choose from to test the pipeline. These include building form datasets with a mixture of LOD1 and LOD2 form data, such as Zurich (Zurich 2020), Hamburg (State Office for Geoinformation and Surveying 2017), and Singapore (Urban Analytics Lab, National University of Singapore 2019). Although CityGML has LOD2 building forms for certain European cities like Berlin (Senate Department for Economic Affairs, Energy and Public Enterprises 2020) and Helsinki (City of Helsinki 2017), their coverage is limited to specific districts such as the central business areas. Among Northern American cities, the New York City dataset (Department of City Planning 2018) contains LOD2 3D data of building forms in the 3DM format, but does not guarantee watertightness. We, therefore, chose the Montreal dataset (Datopian 2016), as it provides numerous LOD2 watertight building forms in the 3DM format. The Montreal dataset is obtained from Montreal Open Data (City of Montreal 2018), a platform that provides access to various urban data collected and processed for internal operations for reuse by its citizens for different purposes. We downloaded 73 3DM files containing over 70,000 building forms in six neighborhoods, namely, Côte-des-Neiges–Notre-Dame-de-Grâce, Outremont, Plateau Mont-Royal, Sud-Ouest, Verdun, and Ville-Marie (Figure 2).

On average, each 3DM file contains approximately 980 building forms, each of which consists of a footprint, walls, and roofs represented as polygon meshes. On average, each 3DM file contains approximately 980 building forms, each of which consists of a footprint, walls, and roofs represented as polygon meshes. To extract the polygon meshes of individual building forms from each file, we developed an algorithm using Grasshopper. The algorithm iterates through each footprint, detects the corresponding walls and roofs, and joins them into a new polygon mesh.

During this process, we constructed two types of polygon meshes: orientation-embedded and normalized. In the case of orientation-embedded meshes, we preserved the original orientation of the buildings as present in the source data. Conversely, for normalized meshes, we designated the longest side of each building footprint's maximum bounding box as the principal axis and aligned the building's orientation with this axis. Consequently, all building forms were aligned to either the 0-degree or 90-degree axis, effectively neutralizing orientation information. Upon normalizing all

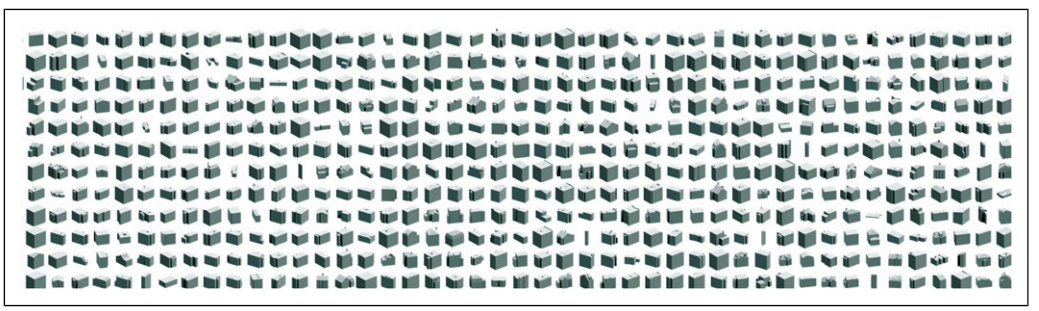

Figure 2. Samples of Montreal building form dataset.

building positions and orientations to the origin and either the X or Y axis, the algorithm proceeded to save each polygon mesh in the OBJ file format. Employing this methodology, we successfully procured two distinct sets, each comprising 72,165 OBJ files.

Table 1 gives an estimated data size of a converted OBJ file as a voxel file and the size of the dataset at different resolutions. Higher resolution voxel files can represent greater detail in the building forms but at the cost of increased file size and exponentially increasing processing time. To balance level of detail against the size of the dataset, we traded off detail for manageable file sizes and consequently chose a resolution of 32. During the conversion process, we also filtered out 3,853 broken, open, or overlapping meshes, which account for 5.3% of the entire dataset.

During the conversion, we utilized signed distance functions (SDF) to continuously represent voxel space. SDF encodes "the point's distance to its closest surface point, where the sign indicates whether the point lies inside (−) or outside (+) the object" (Kleineberg et al. 2020). Voxels generally refers to a set of grid points with discretized Boolean values. When a form occupies a voxel point, its value is 1; and 0, otherwise. However, this binary discretization tends to lose form detail and lowers prediction accuracy in a neural network. By contrast, SDF-based voxels construct continuous space of representation based on distance, increase accuracy of voxel prediction, and compensate for the lack of details from the limited number of voxel grids (Park et al., 2019).

Each building in our dataset is represented by an SDF-based voxel file at a resolution of 32, containing 32,768 floats arranged in a 3D grid. The size of a single SDF-based voxel file is 129 KB, and the entire dataset consists of 68,312 voxel files. The total size of the entire dataset is approximately 8.59 GB.

Model

In this experiment, we used a hybrid generative model of Generative Adversarial Network, GAN (Goodfellow et al., 2014) and Variational Auto-Encoder, VAE (Kingma and Welling 2019) for better performance and representation of morphological features.

A GAN consists of two networks: the generator and discriminator. The generator maps the captured features of the training data to data space—generating a random data sample—while the discriminator distinguishes whether the input sample is real or fake data by evaluating how similar the sample is to real data. According to Larsen et al. 2016, "the GAN objective is to find the binary classifier that gives the best possible discrimination between true and generated data and simultaneously encouraging the generator to fit the true data distribution."

A VAE also consists of two networks: the encoder and decoder. The encoder compresses the input sample to the reduced dimension of data space, the so-called latent space. The decoder reconstructs the encoded sample to resemble the original sample as closely as possible. Training VAE reduces the error between this reconstruction and the original sample by capturing the data distribution in specific dimensions.

Although both GAN and VAE can be used to generate synthetic data, GAN is generally better at producing high-quality samples with fine detail. However, as GAN uses a random sampling process

Table 1. Dataset size estimation by resolutions.

Resolution	Size of a Voxel File	Size of the Dataset
8	3 KB	0.2 GB
16	17 KB	1.2 GB
32	129 KB	9.3 GB
64	1025 KB	73.9 GB

in the generator, it is harder to construct a latent space where a given data can be distributed by their encoded features. By contrast, VAEs can construct a meaningful latent space using their encoder, which captures important features from the input data.

Taking the advantage of both GAN and VAE, we employed a hybrid model—IntroVAE (Huang et al., 2018)—which consists of an encoder and generator. Capturing data features, the encoder compresses a given data sample and expresses it as a latent vector. Like a decoder, the generator reconstructs the latent vector back to the original data, but it also has a random sampling process as in the GAN generator. The sampling process can diversify training of the encoder, increasing detail in reconstruction. In addition to performing adversarial learning like GANs, the inference and generator models are jointly trained for the given training data to preserve the advantages of VAEs (Huang et al., 2018), such as not having the need to train a heavy model, the discriminator. Training IntroVAE decreases joint loss () between the original, reconstructed, and randomly generated samples.

We adapted the IntroVAE architecture, which was originally designed for image data, to a 3D voxel version and conducted a small experiment. Our goal was to compare the learning performance of building forms between a simple VAE and the adapted IntroVAE. Both are trained with the same building form dataset and hyperparameters. Five hundred randomly selected buildings are input to both networks and reconstructed through the encoding and decoding process. We measured their model errors—the occupancy difference in the voxel points between the reconstructed building forms and their ground truth. On average, the error of simple VAE was about 37.45%, while that of IntroVAE was approximately 26.05%. Figure 3 demonstrates that reconstructions from IntroVAE are more similar to the ground truth than those from simple VAE. Moreover, we observe that IntroVAE is capable of capturing feature details such as slope of the roof and extrusion of a chimney. This experiment indicates that IntroVAE exhibits better performance in capturing and generating the morphological features of the building form dataset than simple VAE.

Training

For each dataset, the IntroVAE model was trained on a computer with the following specification: 'Intel Core i9-10,900X @ 3.70 GHz', 256 GB memory, and RTX A6000 graphic processing units. It took approximately 28 h to train the data for 2000 epochs. The hidden dimension, or latent space

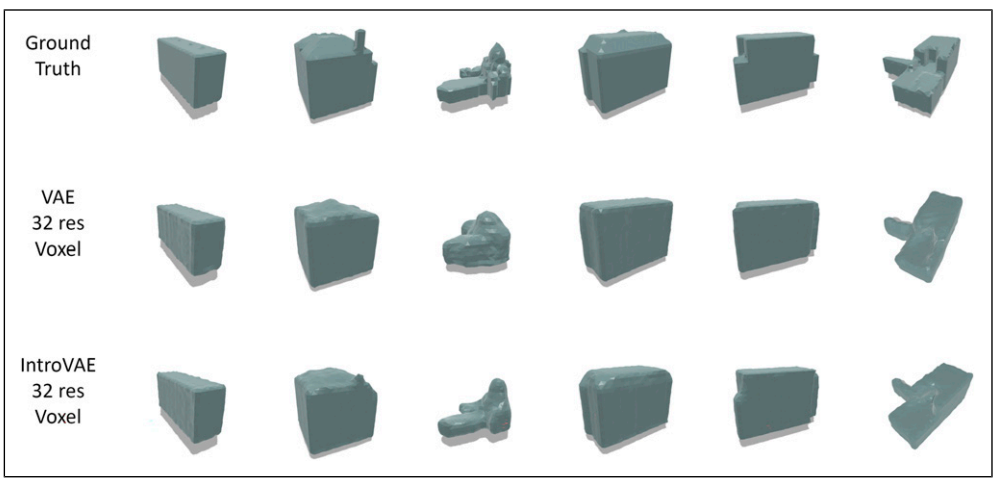

Figure 3. Quality comparison of capturing and reconstructing morphological features between simple VAE and IntroVAE.

dimension, was 256. The model complexity multiplier value was 48, and the batch size was 512. A weight of 0.1 was assigned to the margin loss. The values of alpha and beta were 10 and 0.25, respectively. ADAM (Kingma and Ba 2015) optimizers were used for both the encoder and decoder. The learning rate of the encoder and decoder was 3×10^{-6} until the 550 epoch, 5×10^{-7} from the 550 to 1500 epoch, and 1×10^{-7} from the 1500 to 2000 epoch.

Figure 4 illustrates the quality of the reconstructed building form with respect to the changes in training and testing losses over epochs. Lower losses indicate more precisely captured morphological features. After approximately 80 epochs, the losses began to converge to "a stable stage in which their values fluctuated slightly around a balance line" (Huang et al., 2018), capturing the schematic form. Between 80 and 500 epochs, the loss gradually decreased, capturing intermediate detail. After reducing the learning rate at 550 epochs, the losses decrease slightly, resulting in more accurate detail.

For both datasets, the final training and test losses are approximately 0.19 and 0.27, respectively. The final losses are slightly higher than the expected values from general deep neural network training, which is usually between 5% and 10%. However, when taking into account losses from similar training problems in 3D object research (Wu et al., 2016), the final losses fall within the acceptable range of 9%–25%.

Distinguishing type

After training, morphological features are encoded in each building form dataset, resulting in latent vectors. The shape of the latent vectors is the same as the amount of data by the size of the hidden dimension (68,312 × 256). An encoded vector preserves the morphological features of a building within the reduced data dimension. In other words, each building is represented by their feature vectors of 256 floats in the latent space. We employ t-SNE (t-distributed stochastic neighbor embedding) (Maaten and Hinton 2008) to visualize the latent vectors by reducing their dimensionality.

T-SNE requires four distinct hyperparameters: the number of components, perplexity, learning rate, and iterations. These terms are defined as follows. Number of components refers to the

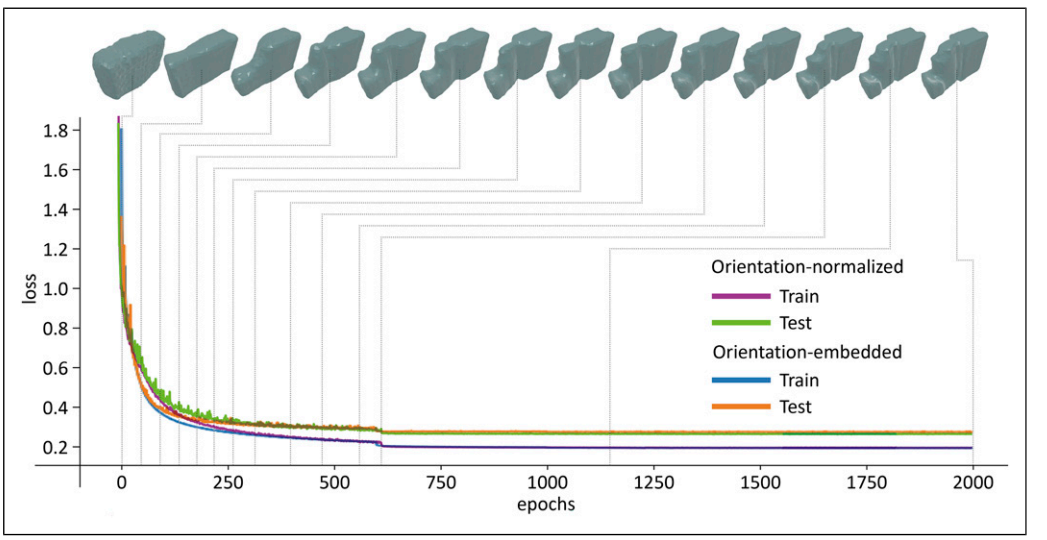

Figure 4. The changes of losses and reconstruction quality during training process.

dimension of the embedded space. Iterations represent the maximum number of iterations for optimization. (Maaten and Hinton 2008) define perplexity as the "smooth measure of the effective number of neighbors." According to them, larger datasets usually require a higher perplexity, and typical values range between 5 and 50. A larger learning rate generates a more spherical distribution of the data embeddings. If the learning rate is too low, the distribution appears as a dense cloud. For the orientation-embedded dataset, we applied the following t-SNE settings: two components, 3,000 iterations, a perplexity of 35, and a learning rate of 50. In contrast, for the orientation-normalized dataset, while maintaining the same number of components, iterations, and learning rate, we adjusted the perplexity to 30 to account for the standardized orientation data.

All the building forms in two datasets are located and represented as data points in a 2D embedding space using t-SNE. The position of each point indicates its morphological characteristics, and the distance between points represents the degree of morphological similarity; the shorter the distance, the more similar are the building forms. By identifying and clustering core samples of high density from the data points using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996), we identified core samples of high density from these data points and subsequently clustered them. This method allowed us to distinguish 20 distinct types in both the orientation-embedded and orientation-normalized datasets. Remapping the data cluster onto map space reveals the distribution of each type (see Figure 5, Figure S1, and Figure S2).

To analyze the characteristics of building form types, three different descriptors of building forms were developed for each type: a distribution map, principal axes, and a representative form (see Figures 6 and 7.) The distribution map is a geo-spatial descriptor, displaying the pattern of how buildings in a type are distributed geographically. The principal axes are the top 50 influential axes on changes in building shape among the 256 latent vector axes. We obtain these axes by measuring the changes in building volume while modifying each axis value throughout the entire dataset. The more changes that occur, the more influential is the principal axis. The average of the principal axes of all buildings in a type quantitatively describes the morphological characteristics of the type. The representative form of each type is defined as the building form of the nearest data point to the centroid of its cluster. This form serves as a qualitative descriptor of the morphological characteristics of a type.

Results

Considering the three descriptors, we grouped the 20 types into five categories, which we named as Group A, B, C, D, and E for the orientation-embedded dataset (Figure 6) and Group F, G, H, I, and J (Figure 7). This section will address the analysis of the characteristics of types in each group and illustrate the comparative analysis of the characteristics in different groups.

Group A, consisting of types 07, 17, 18, and 19, is primarily situated in the city's southeastern region, with buildings exhibiting similar sizes and orientations that align with local street networks. While types 07, 17, and 18 generally face south-east or north-west, type 19 structures orient south-west or north-east (see Figure S3). Despite overall similarity in principal axes with Groups B and D, notable variances exist in values between 110 and 130 on these axes. Morphological distinctions among types within Group A are minimal, marked mostly by differences in orientation—about 12, 10, and 96° for comparisons between type 07 and types 17, 18, and 19, respectively. Type 17 stands out due to its steep gable roofs and architectural details, indicating the model's emphasis on orientation over intricate form differences in classification.

Group B, encompassing types 10 through 13, is predominantly found in the city's northern and southern regions, with buildings aligned parallel or perpendicular to the north-to-south street axes that tilt 35–40° eastward. This group's distinct feature is the significant variance in the principal axes between 220 and 255, with the first 10 axes having higher amplitudes than other groups. Within the

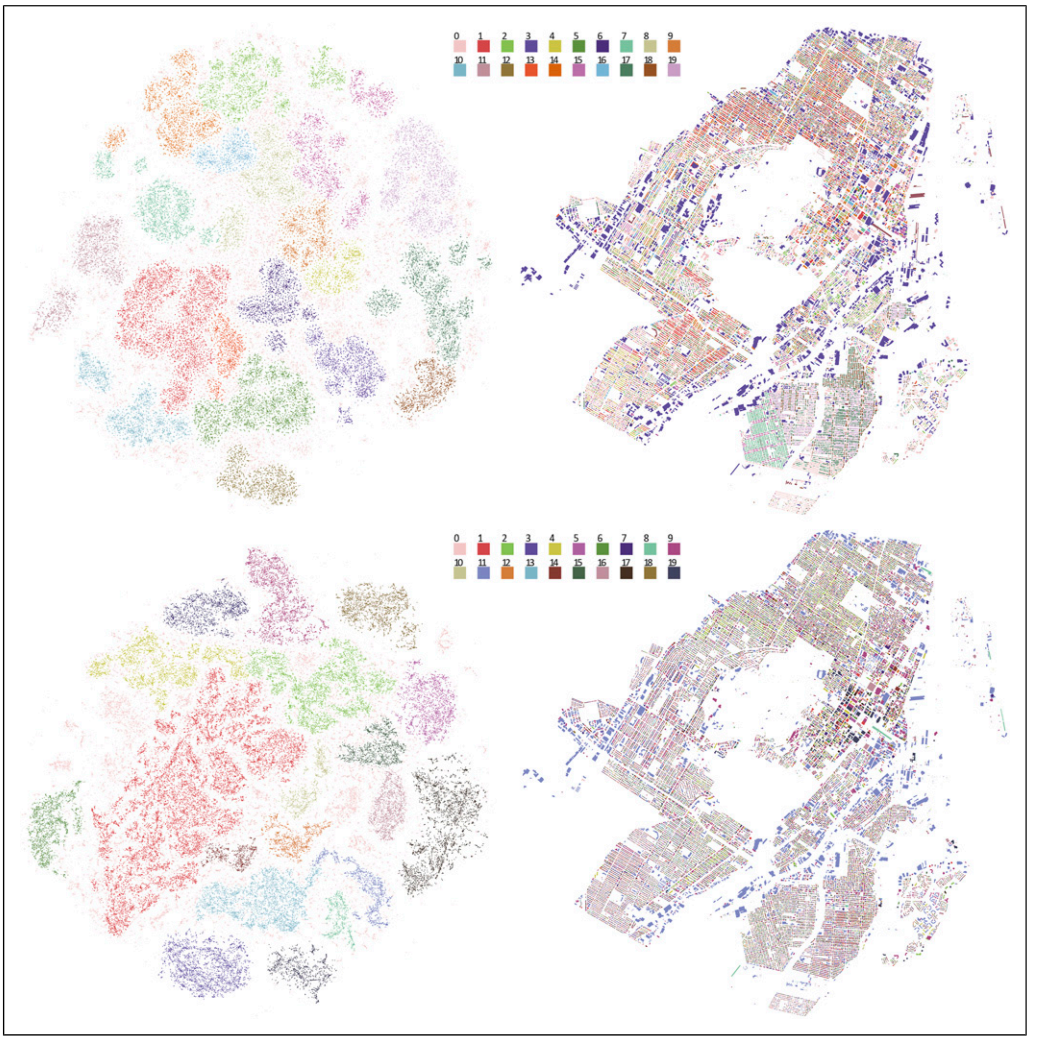

Figure 5. 20 types of orientation-embedded (top) and orientation-normalized (bottom) Montreal building form datasets distribution in data space (left) and map space (right).

group, type 10 features simple cuboid structures akin to row houses, while type 11 has similar forms but with smaller, lower extensions. Type 12 includes dual-cuboid structures, and type 13 presents buildings with L-shaped extrusions, highlighting nuanced morphological variances.

Group C comprises type 02 and 10, which are densely distributed in the central area and have a mild distribution throughout the city. The street axes run from north to south, tilting about 35–40° to the west. The principal axes indicate a trend in the value of the axes with the smallest variance. The noticeable characteristic of principal axes is that the amplitude of the first 50 axes is small. Type 02 buildings are cuboid-shaped row houses, whereas type 10 buildings are more complex with additions or L-shaped footprints.

Group D, the most populous in the orientation-embedded dataset, features types 01, 04, 05, 06, 09, 14, 15, and 16, and is widespread throughout the city, barring the south-east and central regions dominated by groups A and C. This group's principal axes reveal two key aspects: notably smaller

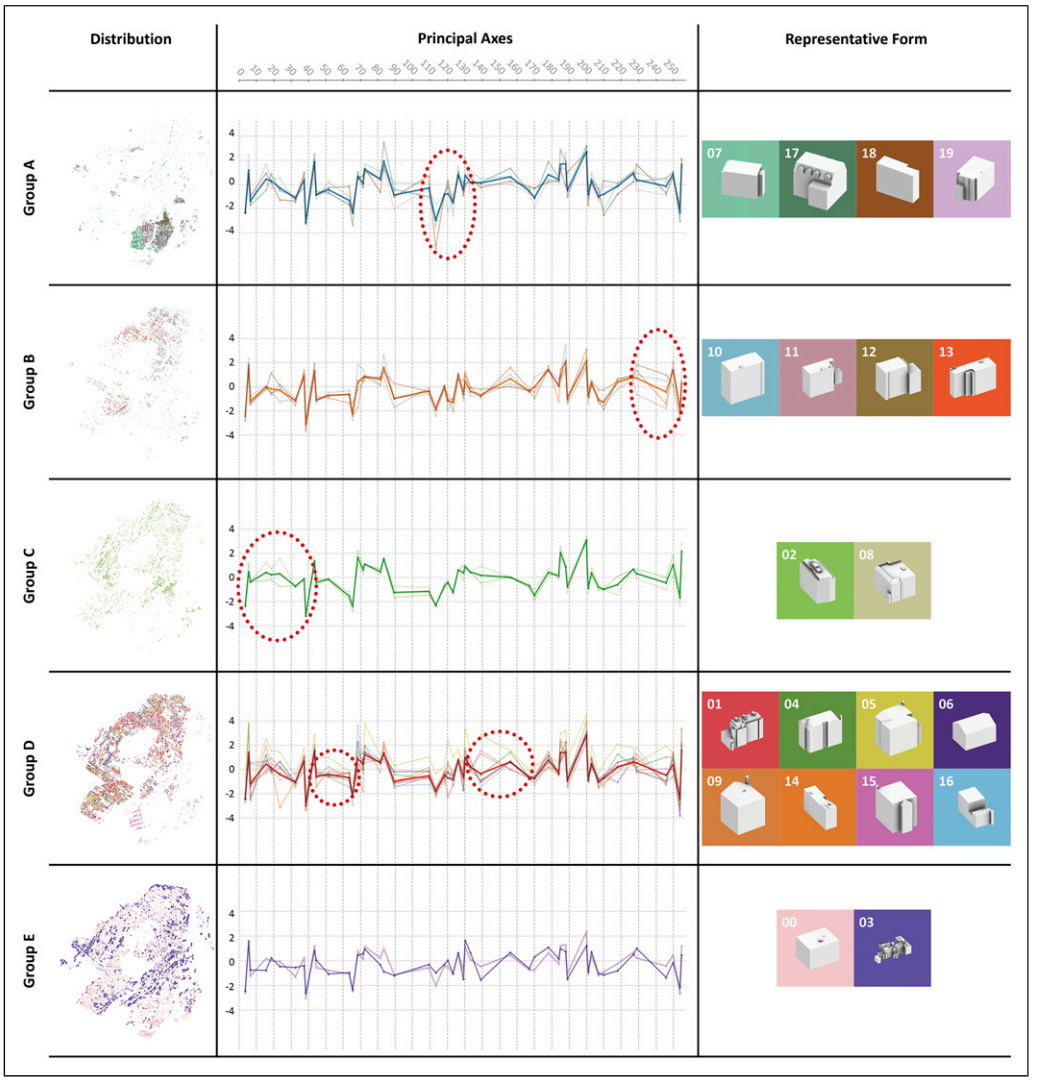

Figure 6. Distribution on map space, principal axes, and representative form of each type by five different groups. The red dotted ellipses indicate the prominent parts of the principal axes. Orientation-embedded dataset.

values between 40 and 70 and high variance between 130 and 170. The types in Group D showcase diverse building elements, from additions and footprints to mass configurations and roof styles, with examples including type 01's flat-roofed cuboids with detailed textures, type 05's mansard roofs with a cubic ratio, and type 16's rectangular forms with orthogonal irregularities.

Group E comprises types 00 and 03, which exhibit miscellaneous characteristics from the other four groups. As this group is not categorized based on a specific morphological tendency, there is a large variance observed throughout the axes. Type 00, indicated by a light pink color, includes unclustered buildings. Type 03 displays monumental building forms with large and complex

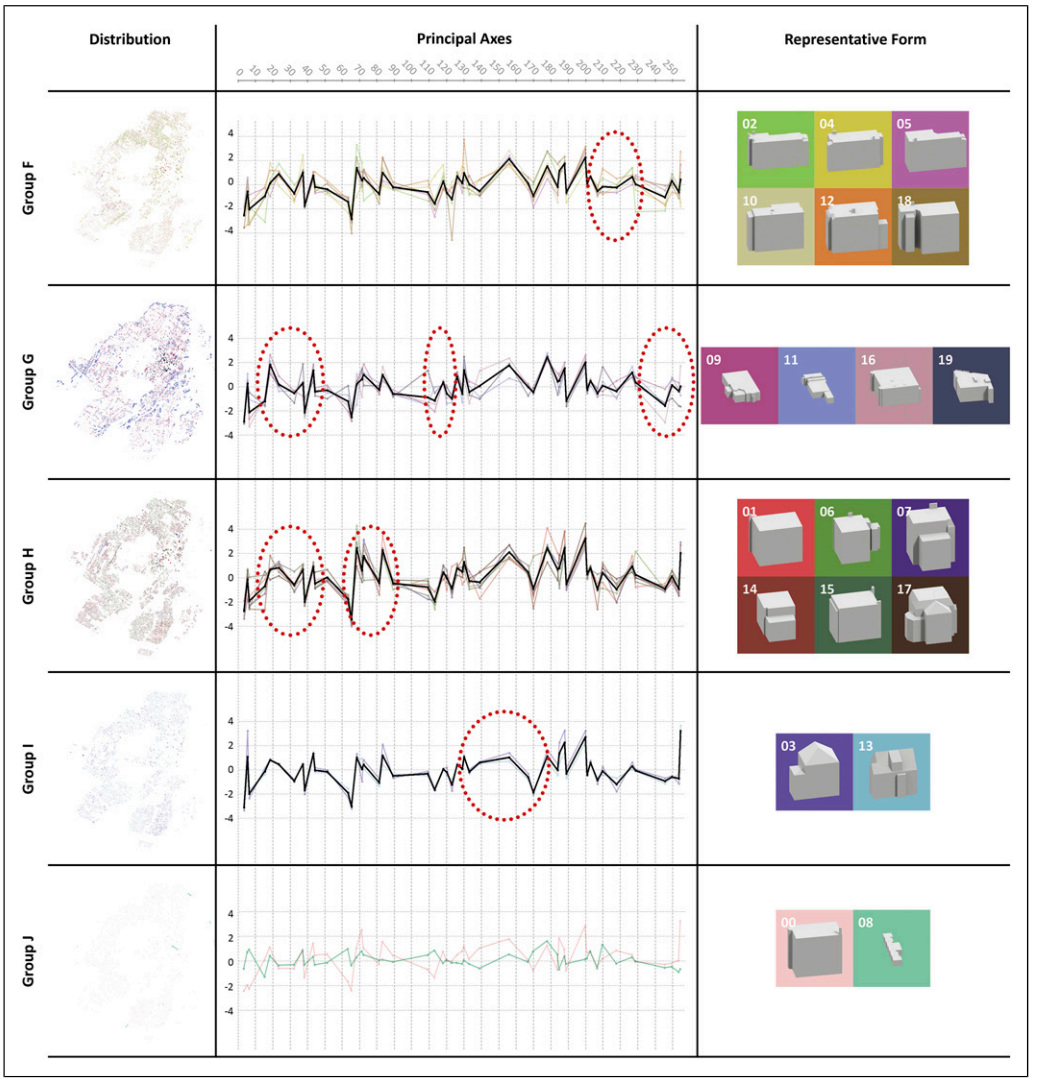

Figure 7. Distribution on map space, principal axes, and representative form of each type by five different groups. The red dotted ellipses indicate the prominent parts of the principal axes. Orientation-normalized dataset.

morphological configurations, such as museums, churches, and community centers. The types in this group are distributed throughout the entire city, with diverse shapes and orientations.

In general, the types appeared to be primarily influenced by the orientation of the building form data. Nonetheless, through a comparative analysis of the morphological characteristics of building types concentrated in certain areas, it is observed that the deep neural network is capable of distinguishing between various interpretable morphological features of buildings, such as the shape of the roof and composition of volumes, over the dominance of orientation in the form data.

Group F consists of types 02, 04, 05, 10, 12, and 18, predominantly situated in the city's northern and southern regions. These types share a commonality in their elongated rectangular form and similar building dimensions, with types 02, 04, 05, and 10 exhibiting comparable width and depth.

In contrast, types 12 and 18 are characterized by a slightly shorter depth. Notably, types 02, 05, 10, and 18 possess L-shaped or near-rectangular footprints, while types 04 and 12 present rectangular forms with attachments and significant corrugation. Though the principal axes of this group bear resemblance to those in Group G, there is substantial variation in the axes values between 200 and 230.

Group G includes types 09, 11, 16, and 19, scattered across the city but notably concentrated near Downtown. These types are significant for their monumental forms with extensive and intricate configurations, akin to museums, churches, and large commercial buildings. Specifically, type 11 stands out due to its considerably larger scale and vertically elongated structures, whereas types 09, 16, and 19 encompass primarily square-shaped monumental structures with proportional width and depth. Three principal characteristics are discernible in this group's axes: a large variance from 10 to 50, significant variance from 110 to 130, and a markedly small variance from 230 to 255.

Group H, the most populous in the orientation-normalized dataset, incorporates types 01, 06, 07, 14, 15, and 17, spread throughout the city. Most buildings in this group showcase square-shaped footprints, with types 07, 14, and 17 exhibiting more attachments and corrugations compared to types 01, 06, and 15. The principal axes reveal two prominent characteristics: unique patterns with relatively small variance in the values from 10 to 50 and a large variance from 60 to 90.

Group I encompasses types 03 and 13, dispersed citywide. Although composed similarly to Group H with square-shaped footprints, the distinction lies in their roofing, with this group featuring gable or hip roofs as opposed to the flat roofs prevalent in Group H. The principal axes display a small variance between 130 and 180.

Group J, mirroring Group E, presents diverse characteristics not specific to any morphological category, leading to broad variance across the axes. Similar to type 00 in Group E, represented by a light pink hue, type 08 manifests monumental forms, significantly large and linear in footprint shape. The distribution of the principal axes is distinct, characterized by lower values and reduced variance, presenting patterns divergent from other types.

In summary, the orientation-normalized dataset indicates that these types are predominantly influenced by aspects related to ratio, scale, roof shape, and attachments, rather than orientation. The deep neural network is effective in identifying detailed morphological configurations without the substantial influence of building orientation. Moreover, the distribution of types generated without orientation exhibits no specific regional characteristics but a citywide dispersion. While unique distribution features were common in the orientation-embedded dataset, the orientation-normalized dataset reveals a quite consistent distribution across the city, implying the model's enhanced capacity to analyze building forms irrespective of urban location in the absence of orientation.

Comparing the form classification between orientation-embedded and orientation-normalized datasets is crucial in understanding the significance of orientation in 3D building form dataset curation and training using deep learning. Firstly, Group B and Group F can be examined comparatively due to their distributional similarities. Both groups are primarily located in the northern and southern parts of the city, and a majority of buildings in these groups have been observed to possess rectangular-shaped footprints. However, Group F encompasses a broader category, suggesting that Group B could be interpreted as a subset possessing specific morphological traits of Group F. Group B consists of buildings with a distinctive orientation facing southwest or northeast, a feature prevalent in structures located in the northern and southern parts of the city. This distribution can also be observed similarly in Group F, even when orientation is excluded, indicating that while orientation-based classifications vary by region, orientation does not always function as an overriding feature among other morphological traits.

Furthermore, high similarity in distribution between Group G and type 03 in Group E, both representing types with monumental buildings, warrants a comparative analysis. Type 03 in Group E encompasses all monumental types not included in the other 19 types, differing from Group G. All

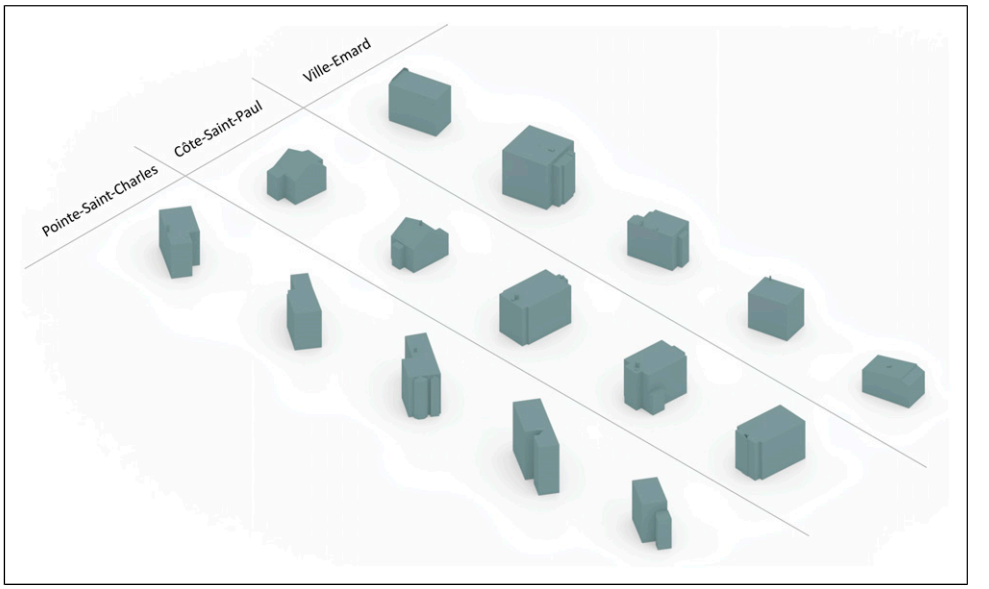

Figure 8. Visual comparison of randomly sampled building forms in Ville-Emard, Côte-Saint-Paul, and Pointe-Saint-Charles.

types in Group G are monumental, but they are classified differently based on proportions, size, and complexity of forms—see Figure S4 for details. Therefore, Group G can be seen as a more nuanced representation of monumental buildings compared to type 03 in Group E.

Through these cases, it is evident that when orientation is normalized, deep learning models can potentially learn more form-related features, facilitating a more nuanced classification of building forms. This suggests that features associated with orientation had a stronger influence on classification than those affecting shape, such as ratio, footprint shape, and height. However, the fact that the deep learning model distinguishes nearly all monumental buildings with various orientations in type 03—albeit not in a nuanced manner—indicates that the model does not always prioritize orientation in classification.

Despite these exceptional circumstances, it is recognized that deep learning acknowledges the role of orientation as a significant feature in classification. Orientation transcends mere building directionality, potentially functioning as a feature closely related to the terrain and urban formation patterns of the region. For instance, type 07 predominantly represents most buildings in Ville-Émard (VE), type 19 those in Côte-Saint-Paul (CSP), and type 02 in Côte-Saint-Paul (CSP) (Figure 8). These buildings maintain orientations parallel or perpendicular to these neighborhoods' street network, despite various morphological variations, confirming the consistency of orientation. This reflects the unique orientations inherent to each area being embedded in the building form data, effectively serving as morphological features distinguishing types. Thus, it is demonstrated that orientation can be trained by deep learning models as an element of urban locality.

Conclusion

In this research, we introduce an inductive approach for the classification of building forms in a city, delving into the implication of building orientation as a feature of urban locality in a large dataset of building forms. By referencing how other morphological studies handle large form data, we propose a pipeline for morphological analysis of building forms using 3D deep neural networks. The

pipeline includes the process of constructing a 3D voxel-based building form dataset, training a hybrid model of GAN and VAE for 3D voxels, encoding morphological features of the dataset, uncovering major types by clustering the features, and examining characteristics of the types according to their distributions on map space, principal axes, and representative forms.

For our case study, we compiled two extensive datasets of building forms in Montreal, characterized as orientation-embedded and orientation-normalized. The application of these datasets within our proposed pipeline enabled the discernment of 20 building form types, aggregated into five categories. We analyzed the morphological attributes inherent to each category and executed a comparative scrutiny amongst them.

Through a juxtaposed evaluation of the disparate building form types derived from the two datasets, several key deductions emerged. Firstly, building orientation, profoundly intertwined with the environmental, social, and cultural contexts of a city, stands out as a crucial determinant in the classification of building forms facilitated by deep learning. Secondly, deep learning is capable of assimilating aspects of urban locality, a feat made possible by the orientation features embedded within the dataset. However, it's critical to note that orientation doesn't always overshadow features associated with morphological traits, such as ratio, scale, and footprint shape, thereby challenging the notion of its universal predominance in form classification. This realization underscores a significant deviation from other morphological studies utilizing deep learning, highlighting the necessity for a nuanced, calibrated approach that involves the intentional inclusion of orientation, specifically tailored to suit the analytical objectives at hand.

In light of these revelations, we contend that in the realm of architectural and urban form analysis through deep learning, the treatment of building orientation's embeddedness can bifurcate into two scenarios: those necessitating inter-city comparisons and those requiring intra-city comparisons. In the former, given the disparities in terrain and cultural contexts across cities, orientation should be normalized, and its implications attenuated to facilitate a coherent comparative analysis of architectural or urban forms. In contrast, the latter hinges on the decision to incorporate regional characteristics; if the analysis focuses solely on shape-specific features, orientation should be expunged, as demonstrated through normalization in this study. Conversely, analyses with a regional emphasis mandate the embeddedness of orientation. For instance, exploring the correlation between solar exposure and building forms in a city unequivocally demands the embeddedness of orientation within the data. Moreover, in cities in East Asia, where there is a pronounced preference for south-facing orientations influencing building forms, the inclusion of orientation information becomes imperative.

Another significant contribution of this study is the presentation of a pipeline for morphological analysis using deep learning. This pipeline enables the measurement of morphological similarities across a vast array of building forms and facilitates their subsequent classification into types. However, the pipeline does not include evaluation methods for the analysis results, raising questions about the reliability of the identified types. As such, for future work, diverse evaluation methods will need to be explored, developed, tested, and integrated into the pipeline, including site surveys, historical reasoning, and interviewing experts on urban development.

Additionally, this case study utilizing the pipeline reveals certain limitations due to the employment of data with a resolution of 32, resulting in blurry forms post-final training, and potentially overlooking detailed features like chimneys and various attachments in the classification process. Consequently, future endeavors should focus on securing access to higher-performance equipment to experiment with higher resolutions and refine the pipeline.

Regarding the data, our division was based on the records provided by the city of Montreal, with buildings delineated according to architectural registries and lots. However, the treatment of many buildings sharing party walls with others was not incorporated into this study, an omission that bears significance given that such structures might be perceived as a single entity within the cityscape.

This issue, crucial in individual architecture, exerts considerable influence on the configuration of urban form data, especially when multiple buildings are present. Therefore, subsequent research needs to explore the variations in deep learning classification contingent upon contiguous architecture.

Despite these limitations, we proposed this analytical pipeline as a preliminary measure for studying phenomena associated with an abundance of three-dimensional architectural and urban forms. We contend that the synergy of this classification is amplified when integrated with additional social, environmental, and cultural data. For instance, an analysis of building forms can be linked to environmental aspects, particularly energy use, and can help identify the characteristics of buildings. The potential of this pipeline grows when the data extends beyond building forms to specifically defined urban forms. For example, if traffic accident data is associated with 3D urban form data, the pipeline can be expanded to analyze the likelihood of traffic accidents occurring based on urban form characteristics. Hence, the pipeline presented in this study serves as a pioneering effort capable of consistently analyzing a vast array of 3D form information, underscoring the necessity for more advanced subsequent research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the This work was supported by the Architectural Design Human Resource Development Project; the Korea Agency for Infrastructure Technology Advancement ($15,000) and Computational Design Laboratory, and the School of Architecture at Carnegie Mellon University ($3,000).

ORCID iD

Jinmo Rhee https://orcid.org/0000-0003-4710-7385

Data Availability Statement

The relevant data and deep learning models used for the paper are in the github repository: https://github.com/leeuack/building_form_analysis_montreal

Supplemental Material

Supplemental material for this article is available online.

References

Biljecki F, Ledoux H and Stoter J (2016) An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems* 59: 25–37. DOI: 10.1016/j.compenvurbsys.2016.04.005.

Boeing G (2017) OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65: 126–139. DOI: 10.1016/j.compenvurbsys.2017.05.004.

Bookstein FL (1992) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511573064.

Cai C, Guo Z, Zhang B, et al. (2021) Urban morphological feature extraction and multi-dimensional similarity analysis based on deep learning approaches. *Sustainability* 13(12): 6859. DOI: 10.3390/su13126859.

City of Helsinki. (2017) "The City Information Model." 2017. https://www.opendata.fi/data/en_GB/dataset/helsingin-3d-kaupunkimalli/resource/577f4286-7162-42e9-8ffe-52632228569e.

City of Montreal (2018) *"Montreal Open Data."* Montreal, Données Ouvertes. https://donnees.montreal.ca/

Datopian (2016) "Bâtiments 3D (Maquette LOD2 Avec Textures) - Dataset." https://donnees.montreal.ca/ville-de-montreal/batiment-3d-2016-maquette-citygml-lod2-avec-textures2

Department of City Planning (2018) "NYC 3D Model." https://www.nyc.gov/site/planning/data-maps/open-data/dwn-nyc-3d-model-download.page

Dinsdale NK, Smith SM, Arya Z, et al. (2021) Learning Patterns of the Ageing Brain in MRI Using Deep Convolutional Networks. *NeuroImage* 224: 117401. DOI: 10.1016/j.neuroimage.2020.117401.

Escamilla-Rivera C, Alejandra Carvajal Quintero M and Capozziello S (2020) A deep learning approach to cosmological dark energy models. *Journal of Cosmology and Astroparticle Physics* 2020(03). DOI: 10.1088/1475-7516/2020/03/008.

Ester M, Kriegel HP, Sander J, et al. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96) - Portland, OR, USA. August 2-4, 1996: 226-232.

Fleischmann M, Feliciotti A and Kerr W (2021) Evolution of urban patterns: urban morphology as an open reproducible data science. *Geographical Analysis* 54(3): 536–558. DOI: 10.1111/gean.12302.

Goodfellow IJ, Pouget-Abadie J, Mehdi Mirza M, et al. (2014) "Generative adversarial nets." In *Advances in Neural Information Processing Systems*. Montreal: Universite de Montreal. https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

Huang H, Li Z, He R, et al. (2018) IntroVAE: introspective variational autoencoders for photographic image Synthesis. *Advances in Neural Information Processing Systems*. New York, USA: Curran Associates, Inc, vol. 31. Available at: https://proceedings.neurips.cc/paper/2018/file/093f65e080a295f8076b1c5722a46aa2-Paper.pdf.

Idgunji S, Ho M, Payne JL, et al. (2021) Deep Neural Networks for Hierarchical Taxonomic Fossil Classification of Carbonate Skeletal Grains. EGU21-16394.

Kalinin AA, Allyn-Feuer A, Ade A, et al. (2018) 3D cell nuclear morphology: microscopy imaging dataset and voxel-based morphometry classification results Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA. 18-22 June 2018: 2272–2280. https://openaccess.thecvf.com/content_cvpr_2018_workshops/w44/html/Kalinin_3D_Cell_Nuclear_CVPR_2018_paper.html

Kingma DP and Jimmy B (2015) Adam: a method for stochastic optimization The Proceedings of ICLR 2015. https://arxiv.org/abs/1412.6980

Kingma DP and Max W (2019) *An Introduction to Variational Autoencoders*. Hanover: Now Publishers. https://ieeexplore.ieee.org/document/9051780

Kleineberg M, Fey M and Frank W (2020) Adversarial generation of continuous implicit shape representations." In The Eurographics Association. *May* 25-29: 2020. doi: 10.2312/egs.20201013

LarsenLindbo AB, Soren KS, Larochelle H, et al. (2016) Autoencoding beyond pixels using a learned similarity metric. Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume. New York, NY, USA. June 19-24, 2016: 1558–1566.

Li P and Zhu W (2020) Clustering and morphological analysis of campus context Distributed Proximities. https://papers.cumincad.org/cgi-bin/works/paper/acadia20_170

Maaten Lvan der and Hinton G (2008) Viualizing data using T-SNE. *Journal of Machine Learning Research* 9: 2579–2605.

Marcus LF (1990) Traditional morphometrics. *Proceedings of the Michigan Morphometric Workshop*: Ann Arbor MI: The University of Michigan Museum of Zoology: Rohlf and F. L. Bookstein: 77–122.

Meeran A and Joyce SC (2020) Machine learning for comparative urban planning at scale: an aviation case study Distributed Proximities. ACADIA. https://papers.cumincad.org/cgi-bin/works/paper/acadia20_178

Miguel Jde, Eugenia Villafañe M, Piškorec L, et al. (2019) Deep form finding: using variational autoencoders for deep form FInding of structural typologies. *Architecture in the Age of the 4th Industrial Revolution*: Porto, Portugal: University of Porto: 71–80. https://papers.cumincad.org/data/works/att/ecaadesigradi2019_514.pdf

Moosavi V (2017) Urban morphology meets deep learning: exploring urban forms in one million cities, town and villages across the planet. arXiv:1709.02939 [Cs]. https://arxiv.org/abs/1709.02939

Newton D (2019) Generative deep learning in architectural design. *Technology Architecture Design* 3(2): 176–189. DOI: 10.1080/24751448.2019.1640536

Ostwald M and Vaughan J (2012) *Significant Lines: Representing Architecture for Computational Analysis*. Proceedings of the 45th Annual Conference of the Australian and New Zealand Architectural Science Association. November 14-16, 2012. Gold Coast, Australia.

Park JJ, Florence P, Straub J, et al. (2019) DeepSDF: learning continuous signed distance functions for shape representation. CVPR 2019. https://arxiv.org/abs/1901.05103

Quan SJ (2022) Urban-GAN: an artificial intelligence-aided computation system for plural urban design. *Environment and Planning B: Urban Analytics and City Science* 49(9): 2500–2515. DOI: 10.1177/23998083221100550.

Rebsamen M, Suter Y, Roland W, et al. (2020) Brain morphometry estimation: from hours to Seconds using deep learning. *Frontiers in Neurology* 11: 244. DOI: 10.3389/fneur.2020.00244.

Ren Y and Zheng H (2020) "The Spire of AI - voxel-based 3D neural style transfer." In *Reinwardtia: Anthropocene, Design in the Age of Humans - Proceedings of the 25th CAADRIA Conference*, Bangkok, Thailand. Chulalongkorn University. https://papers.cumincad.org/cgi-bin/works/paper/caadria2020_091

Rhee J, Cardoso Llach D and Krishnamurti R (2019) Context-rich urban analysis using machine learning - a case study in Pittsburgh In *Architecture in the Age of the 4th Industrial Revolution*, Porto, Portugal: University of Porto. https://papers.cumincad.org/cgi-bin/works/paper/ecaadesigradi2019_550

Schirmer P and Kay A (2015) A multiscale classi;cation of urban morphology. *Journal of Transport and Land Use* 9. DOI: 10.5198/jtlu.2015.667.

Senate Department for Economic Affairs (2020) Energy and Public Enterprises. Berlin 3D." 2020. https://www.opendata.fi/data/en_GB/dataset/helsingin-3d-kaupunkimalli/resource/577f4286-7162-42e9-8ffe-52632228569e

State Office for Geoinformation and Surveying (2017) "3D-Stadtmodell LoD2-DE Hamburg." https://suche.transparenz.hamburg.de/dataset/3d-stadtmodell-lod2-de-hamburg2

Steadman P, Bruhns H, Holtier S, et al. (2000) A classification of built forms. *Environment and Planning B: Planning and Design* 27: 73–91. DOI: 10.1068/bst7.

Urban Analytics Lab, National University of Singapore. (2019) "3D City Model of Singapore Public Housing Buildings." https://github.com/ualsg/hdb3d-data

Webster M and David Sheets H (2010) A practical introduction to landmark-based geometric morphometrics. *Quant Meth Paleobiol* 16: 163–188.

Wu J, Zhang C, Xue T, et al. (2016) Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. Red Hook, NY, USA: Curran Associates Inc.

Yao J (2021) Generative design method of building group - based on generative adversarial network and genetic algorithm. In: Globa A, van Ameijde J, Fingrut A, et al. (eds) PROJECTIONS - Proceedings of the 26th CAADRIA Conference - Volume 1, the Chinese University of Hong Kong and Online, Hong Kong, 29 March - 1 April 2021. 61–70. https://papers.cumincad.org/cgi-bin/works/paper/caadria2021_156

Zhang F, Tian S, Chen S, et al. (2019) Voxel-based morphometry: improving the diagnosis of Alzheimer's disease based on an extreme learning machine method from the ADNI cohort. *Neuroscience* 414: 273–279. DOI: 10.1016/j.neuroscience.2019.05.014.

Zurich, Open Data Zurich (2020) "Open Data Zürich - Stadt Zürich." https://data.stadt-zuerich.ch/dataset?q=&tags=3d-stadtmodell

Jinmo Rhee (Corresponding Author) School of Architecture Planning + Landscape, University of Calgary jinmo.rhee@ucalgary.ca/Jinmo Rhee, an assistant professor at School of Architecture Planning + Landscape at University of Calgary, is a computational designer, architect, and design scholar, delving into the transformative realm of artificial intelligence technologies within architectural design and built environment research. He holds a Master of Science in Computational Design from Carnegie Mellon University and is currently completing his Ph.D. within the same institution. His research centers on understanding architectural and urban forms as conduits for socio-cultural phenomena, employing a data scientific framework to reinterpret design knowledge with a focus on uncovering novel and creative design methods.

Ramesh Krishnamurti School of Architecture, Carnegie Mellon University ramesh@cmu.edu Ramesh Krishnamurti is a Professor Emeritus at the School of Architecture. He has a multidisciplinary research background, focussing on the formal, semantic and algorithmic aspects of generative construction and the development of design as computation via highly coupled parallel explorations of form and description. He is perhaps best known for his work in shape grammars and on spatial algorithms. He has degrees in Electrical Engineering and Computer Science and earned postgraduate degrees in Systems Design at the University of Waterloo. His research was supported by major sponsors including SERC, NSF, DARPA, CERL, and Autodesk.