This is the Accepted Manuscript version of an article published by Elsevier in the journal *Computers, Environment and Urban Systems* in 2024, which is available at:
https://doi.org/10.1016/j.compenvurbsys.2024.102076
Cite as: Wang J, Huang W, Biljecki F (2024): Learning visual features from figure-ground maps for urban morphology discovery. *Computers, Environment and Urban Systems*, 109: 102076.

Learning visual features from figure-ground maps for urban morphology discovery

Jing Wang[a], Weiming Huang[b], Filip Biljecki[a,c,*]

[a]*Department of Architecture, National University of Singapore, Singapore*
[b]*School of Computer Science and Engineering, Nanyang Technological University, Singapore*
[c]*Department of Real Estate, National University of Singapore, Singapore*

Abstract

Most studies of urban morphology rely on morphometrics, such as building area and street length. However, these methods often fall short in capturing visual patterns that carry abundant information about the configuration of urban elements and how they interact spatially. In this study, we introduce a novel method for learning morphology features based on figure-ground maps, which leverages recent developments in computer vision. Our method facilitates discovering and comparing urban form types in a fully unsupervised manner. Specifically, we examine building fabrics by 1 km patches. A visual representation learning model (SimCLR) casts each patch into a latent embedding space where similar patches are clustered while dissimilar patches are dispelled, thus generating morphology representations that entail the layout of building groups. The learned morphology features are tested in urban form typology clustering and comparison tasks in four diverse cities: Singapore, San Francisco, Barcelona, and Amsterdam, with data sourced from OpenStreetMap. Clustering results show effective identification of typical urban morphology types corresponding to urban functions and historical developments. Further analyses based on the representations reveal inner- and cross-city morphological homogeneity relating to socio-economic drivers. We conclude that this method is a promising alternative for effectively describing urban patterns in morphology analysis.

Keywords: Urban form, Deep learning, Computer vision, Unsupervised learning, Representation learning, Cluster analysis

1. Introduction

Urban physical spaces are shaped by intertwining historical, socio-economic and geographical contexts. The study of morphology is a prevalent instrument for analysing urban physical spaces, providing researchers and planners with a comprehensible interpretation of the complexity and locality inherent in cities (Kropf, 2018; Boeing, 2021).

A traditional paradigm of studying urban form involves the interpretation of figure-ground maps (also known as Nolli maps) — in diagrams illustrating urban spaces, land plots or buildings are depicted as black solid mass (figure) while streets or open spaces are represented as white void (ground). Among the urban elements (street, plot, and building), building figure-ground maps provide the most detailed and accurate representation of how people experience and understand urban spaces (Nasar, 1989). In a figure-ground map, the grains, configurations, and the underlying spatial logic organising these urban elements become distinctly visible. (Rowe and Koetter, 1984). For example, in a seminal urban form study, Jacobs (1993) surveyed a wide array of street types and discussed the features of great streets based on figure-ground maps. Such a method offers a lens to understand the formation, transition, and interaction within city organisation (e.g. top-down structural planning versus bottom-up organic growth fabric) (Moudon, 1997; Batty, 2009; Whitehand et al., 2011), also aligns with the culture of visual expression in urban design and planning (Trancik, 1991; Hebbert, 2016).

With the advent of new geospatial data and tools (Yap et al., 2022), another prominent approach of morphology studies relies on morphometrics, which quantifies the geometry or relationships of urban elements (Berghauser Pont and Haupt, 2005; Bocher et al., 2018; Zhang et al., 2023). Streets, for example, are decomposed into average length, connectivity, density and centrality (Boeing, 2017), whereas buildings are represented by footprint area, height, complexity and so on (Biljecki and Chow, 2022). They are then related to a myriad of phenomena such as spatial vitality, energy use, and travel behaviours, in multi-disciplinary studies(Ye et al., 2018; Berghauser Pont et al., 2019; Choi, 2018; Quan and Li, 2021; Xia et al., 2022; Bansal and Quan, 2022).

The morphometrics approach has gained strong momentum in the current urban research landscape across many disciplines. However, the features exam-

*Corresponding author

Email addresses: wangjing@u.nus.edu (Jing Wang), weiming.huang@ntu.edu.sg (Weiming Huang), filip@nus.edu.sg (Filip Biljecki)

ined by visual perception, which reveal the history and spatial logic of urban space have been overlooked. While some morphological features are quantifiable, there are implicit features that are difficult to quantify. These features, however, can be easily understood by humans through visual interpretation, for example, building shape (e.g. podium, slab), street pattern (e.g. organic or grid), and porosity of building groups. These often overlooked features are critical components in the traditional qualitative urban morphology analysis such as the historico-geographical approach (Conzen, 1960, 2004) and the process typological approach (Cataldai et al., 2002), where the visual inspection by local experts and considerable manual effort are necessary. These approaches rely on such features because of their importance in both the philosophical aspect of urban morphology — offering a precise description of the urban landscape, and the cultural aspect — capturing the sense of identity and spirit of place (Barke, 2018).

In this study, we seek to simulate human visual interpretation of urban figure-ground maps with machine eyes. Advances in computer vision (CV) are highly inspirational in this context — convolutional neural networks (CNN) have led breakthroughs on classifying challenging images (Krizhevsky et al., 2012), recognising handwritten zip code digits (LeCun et al., 1989), and distinguishing local binary patterns (Ojala et al., 2002). Considering there is no consensus in the categorisation of urban morphology, we believe unsupervised learning is more feasible than supervised learning in this task. Unsupervised learning discovers patterns from intrinsic data structures, and is gaining momentum in urban studies (Wang and Biljecki, 2022). In computer vision, there have been endeavours to develop unsupervised visual interpretation methods (i.e. visual representation learning) (Chen et al., 2020; Caron et al., 2021). Such techniques generate representations (vector embeddings) of images, e.g. in a contrastive learning framework. Their essence is that similar images would generate embeddings that are close in the embedding space, while the embeddings of dissimilar images would be kept away, and subsequently a simple clustering can uncover the hidden patterns in the data.

By tailoring such techniques, we propose a novel morphology feature learning[1] method based on figure-ground maps. Through feeding figure-ground maps with a few simple morphology metrics into the proposed method, it is able to encode implicit features and cluster similar urban form patches, facilitating discovering and comparing urban form typologies in a fully unsupervised manner.

[1]In this paper, we use the terms “feature learning” and “representation learning” interchangeably.

Specifically, we first lattice the building figure-ground maps into 1 km x 1km patches, and each patch is enriched with 3 simple morphology metrics in the form of numerical values in image channels. Furthermore, we employ an unsupervised visual representation learning framework to encode the patches as vector representations that carry both the spatial layout of each patch as well as the overall geometric features entailed from the morphology metrics. Lastly, we test the representations in four cities located on different continents: San Francisco, Singapore, Amsterdam, and Barcelona. To accomplish this, we engage in typo-morphology discovery through clustering analysis. Additionally, we conduct an evaluation of the inner- and cross-city homogeneity and heterogeneity in urban form by assessing the compactness and similarities of the clusters.

To the best of our knowledge, this study is a pioneering exploration of figure-ground-based alternatives for representing urban morphology. Compared with morphometrics, the advantages of the proposed method lie in that both the collective spatial layout of building groups and the geometry of individual buildings can be encoded, facilitating subsequent discovery process. Beyond this, in comparison with human visual interpretation, our proposed method is able to process large cross-city datasets at one time, in addition, provides quantitative information that benefits similarity evaluation and other potential downstream analyses such as simulation-based study.

We choose to focus on building layers because they are the most dominant man-made element in cities, and their shapes, configurations, and densities largely reflect socio-economic properties. For example, Salazar Miranda (2020) studies how building configurations mediate immigrant social segregation; Ignatieva and Stewart (2009) find that former colonial cities exhibit remarkable homogeneity in building landscapes. When compared to the street or block layers, the building layer demonstrates a higher level of granularity, diversity, and hybridity. Therefore, building morphology analysis is a meaningful and challenging task that calls for the exploration of alternative methods.

The remainder of the paper is organised as follows: Section 2 reviews the related morphology studies and the advances in visual representation learning. Section 3 provides the details of the methodology, study areas, and data. The urban form typologies discovery and comparison results are presented in Section 4. The paper ends with a discussion in Section 5 and conclusions in Section 6.

2. Background and related work

2.1. Representing urban form for typo-morphology analysis

Finding urban form typologies (or morphological areas) is a way to simplify the complexity of cities. Conventionally, researchers read several layers of town-plan maps and manually outline areas with homogeneous urban forms based on systematic stratification criteria (Conzen, 1960; Oliveira and Yaygın, 2020). With the advent of spatial data, this topic has been widely explored through computational methods. Key components in this regard include (1). Input data; (2). Representation (feature extraction) of urban morphology, and (3). Clustering or classification. In the aspect of input data, there are varying choices that could be used for morphological area analysis, e.g., buildings (footprints) (Zhu et al., 2020; Esch et al., 2022) and road network (Boeing, 2017). In the clustering aspect, k-means is the most common method (Song and Knaap, 2007; Gil et al., 2012; Bobkova et al., 2021). A broad range of techniques, including Hierarchical clustering (Serra et al., 2018; Dibble et al., 2019), Gaussian Mixture Model (Jochem et al., 2021; Li and Quan, 2023), and geographically explicit measures of cluster fit (Wolf et al., 2021), have also been explored.

In contrast to the diversity in methodology in the other two components, studies of extracting features to represent urban morphology concentrate on a single stream of 'urban morphometrics', i.e. a rich list of numeric indicators describing various aspects of urban form. Gil et al. (2012) mine 25 dimensions of block and street, focusing on features such as gross floor area, number of buildings, and solar orientation. Vanderhaegen and Canters (2017) propose a series of new metrics to capture the spatial arrangement of built-up areas from patch, profile, and building perspectives. In recent works, more comprehensive lists of metrics are summarised. Biljecki and Chow (2022) form a list of building indicators, with 86 metrics at the building level, and expand to 354 indicators when aggregated at the zone/grid level. Fleischmann et al. (2022) design a numerical taxonomy of urban form (buildings, streets, and plots) with 74 primary (geometric and configurational) characters and 296 contextual characters, as the foundation to conduct subsequent cluster analysis. The same metrics are coupled with functional data to identify and characterise spatial signatures in Great Britain (Fleischmann and Arribas-Bel, 2022). There is a wealth of papers that apply the same methodology to find homogeneous urban types (Wheeler, 2015; Alexiou et al., 2016; Berghauser Pont et al., 2019; Hecht et al., 2013; Schirmer and Axhausen, 2019; Perez et al., 2020). Arguably, morphometrics is the cornerstone of computational urban morphology analysis.

Fleischmann et al. (2022) analogise this method to early biologists seeking to classify biotic species based on morphological similarity, an approach that is

now replaced by DNA sequencing. The studies listed above reveal a limitation of urban morphometrics: It is challenging to construct an exhaustive list of metrics that captures all implicit features and interactions of urban elements, and with the growing number of indicators, the time needed for data processing increases substantially. A direction that has barely been explored is if there are other ways to represent the complex relations and arrangements of urban elements. Can we extract the "DNA" of urban fabric instead of trying to construct various features from urban form?

2.2. Unsupervised visual representation learning

Representation of data aims to gain discriminative information from raw data. Generally, it has two ways: feature engineering and representation learning (feature learning). Feature engineering relies on human ingenuity and prior knowledge, where much effort is spent in designing data preprocessing pipelines. In contrast, representation learning learns the features that entail the intrinsic characteristics of the data through certain training objectives (usually in a unsupervised manner). In the machine learning community, it is a rapidly growing research area accompanied by remarkable empirical successes (Bengio et al., 2013)

Learning effective visual representations without human supervision is highly relevant in the context of this study. Its major incentive is to diminish the reliance on ground truth labels for training CV models, e.g. convolutional neural networks (CNN) (Hamilton et al., 2018). The common practice in this direction is to learn vector embeddings (representations) carrying the similarities between different images, i.e. similar images' embeddings tend to be close and vice versa. Most visual representation learning methods fall into two strands: generative and contrastive.

Recently, contrastive models relying on contrastive learning in the latent embedding space have gained tremendous momentum in visual representation learning. In this regard, seminal models such as Moco (He et al., 2020), SimCLR (Chen et al., 2020), and SwAV (Caron et al., 2021) are classical methods. Among them, Moco relies on momentum contrast to construct a memory bank; Swav utilises the idea of clustering to learn visual representations; SimCLR is based on interactions between different augmented views of the images, i.e. similar views are pulled together while dissimilar views are pushed away.

The boom of visual representation learning studies in CV has also inspired its utilisation in geospatial and urban contexts. For example, Liu et al. (2018) use contrastive learning for points of interest (POI)-based similar region search, in

which they regard each region like an image that is fed into a CNN. The growing interest in contrastive learning has also fuelled unsupervised learning for remote sensing images. Stojnić and Risojević (2021) summarise the applicability of contrastive learning in remote sensing image classification, and discover that using unsupervised contrastive pre-training on remote sensing images can produce comparable performances to supervised training. Bai et al. (2023) combined visual representation learning with cross-modal contrastive learning for learning remote sensing image representations enriched by POIs. van Strien and Adrienne Grêt-Regamey (2022) encode multi-spectrum remote sensing image tiles by Deep Convolutional Embedded Clustering, and divide them into 45 landscape classes.

In the field of urban morphology, several recent studies have made notable contributions with relevant methods. Moosavi (2017) learns "urban vectors" from extensive street networks through a deep convolutional auto-encoder. Additionally, Cai and Chen (2022) introduce a novel approach utilising a convolutional autoencoder to learn representations from satellite images. Both studies showcase the efficacy of these representations in downstream applications, including urban form clustering analysis. Nonetheless, these studies do not delve into the finest-grained patterns of urban form intricately shaped by individual buildings. While satellite images do capture building features to a certain extent, they are not comparable with building footprint vector data in terms of precision and clarity. Furthermore, their representation learning processes lack the inclusion of contextual attributes (such as street categories) that could contribute to more informative representations. At last, the ongoing advancement of machine learning techniques has led to the emergence of more powerful models, offering potential new avenues.

3. Methodology

3.1. Overview

Figure 1 illustrates the research framework of this study.

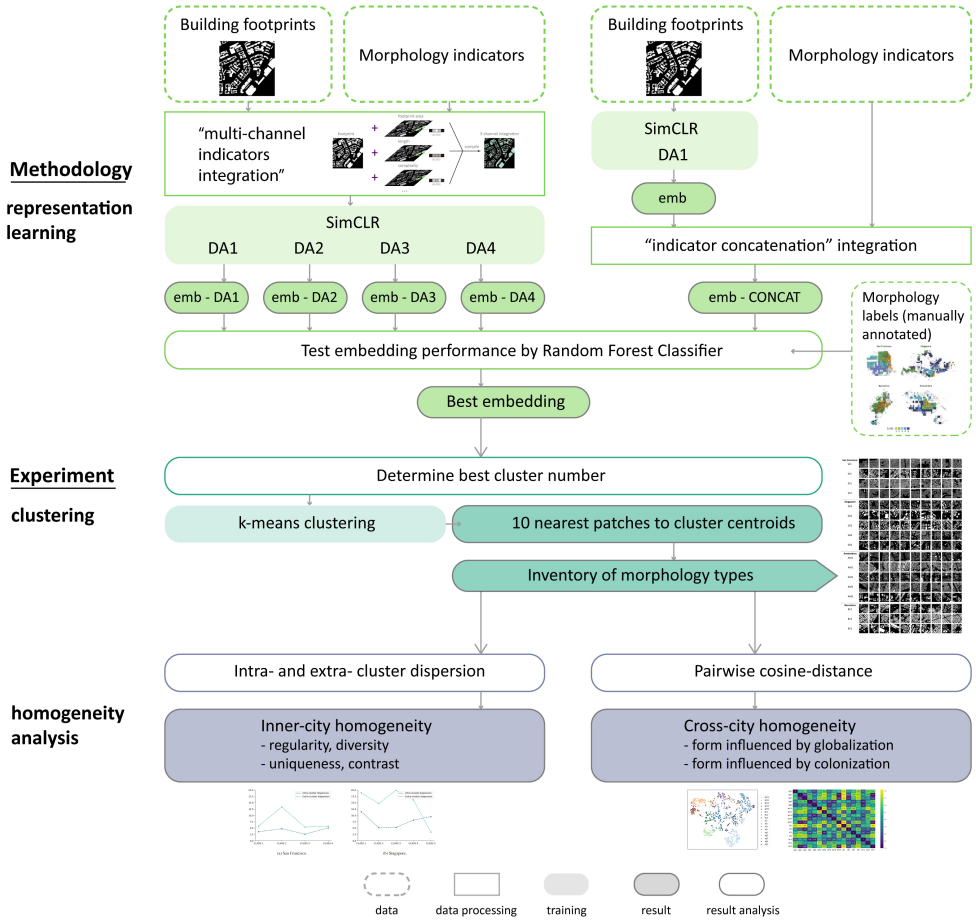


Figure 1: Research framework of our work. Key: DA — data augmentation pipeline; emb — embedding.

The core of our method is morphology representation learning. We adopt an unsupervised framework for learning visual representations — SimCLR (Chen et al., 2020) (elaborated in Section 3.4), to learn representations of building groups based on visual features and a small set of building indicators. Considering the features characterising urban morphology may be different from natural images such as those in ImageNet, we adapt the data argumentation pipeline for our task, and assess the performances based on ground-truth labels, therefore choosing the optimum data augmentation pipeline. The ground truth labels are manually crafted, only for the purpose of evaluating model performance. To incorporate contextual attributes in the learning process, we further develop two means for

integrating building indicators into the training image. The performances of the two integration methods are again evaluated through labels. The best-performing data integration method coupled with the optimum morphology learning model forms our proposed pipeline for morphology representation learning. The sources of our building footprint and morphology data are all publicly available, including OpenStreetMap (OSM) and Global Building Morphology Indicators (GBMI) (Biljecki and Chow, 2022), ensuring the proposed method is transferable to other geographical locations. Representations of the building figure-ground maps (i.e. vector embeddings that allow a system to understand unique characteristics differentiating one pattern from another) are the output of the visual representation learning step.

In the experiment, we utilise a clustering technique (k-means here) to discover similarity patterns based on the representations. We expect that the most representative patches would reside near to cluster centres. In this case, we extract a set of closest patches for each clustering centre, so as to comprehend the prominent morphology types in our study areas. In addition, we conduct various statistical analyses to uncover multi-faceted morphology patterns, e.g. intra- and cross-city morphology homogeneity.

3.2. Study area

Urban morphology patterns have large variations in different cultural and geographical contexts. For testing the effectiveness and transferability of the proposed method, four cities across three continents are selected as study areas: Singapore in Asia, San Francisco in North America, and Amsterdam and Barcelona in Europe. The four cities are typical among cities with similar backgrounds, demonstrating distinctive characteristics, and are morphologically different, providing diversity for our experiments.

San Francisco was first established by Spanish settlers. During the Gold Rush, the city was structurally planned in response to the population surge. The city layout follows an orthogonal grid system, which is similar to the Europe block pattern for bringing people close to jobs (Godfrey, 1997). Moreover, the values of regular land parcels are easy to evaluate thus facilitating land transactions in the market. Such a grid pattern is widely employed in North American cities such as New York City and Toronto (Reps, 1965).

Singapore is a densely populated Asian city-state. To accommodate more than five million residents on limited land, Singapore planned high-rise residential towers dotting the island. In addition, to decongest the living environment and inject greenery into urban life, Singapore was planned in an organic form in contrast to

the grid system (Liu, 2015). As Singapore was historically influenced also by the West, it features numerous 'shophouses' that are similar to some types of European housing.

Amsterdam's morphology is a mirror of its urban development process. The oldest part, the inner city canal ring, was constructed in 1538 and has been preserved to this day. In the 17th century, the city expanded concentrically surrounding the old town. In the 1990s a second large expansion further expanded the city's domain. Today, the city is developing towards the west following a General Extension Plan (Savini et al., 2016). At each stage of development, the urban forms are continuous yet have slight and distinguishable differences.

Barcelona's distinctive block city layout is shaped by Cerdà's Utopian plan in 1855, which is characterised by high egalitarianism. Each block is designed to be identical in size, shape, and building height (Aibar and Bijker, 1997). Correspondingly, parks, shops and housing are distributed evenly throughout the city along with the grid fabric. The block pattern is intertwined with the old medieval city with dense and perpendicular grid streets.

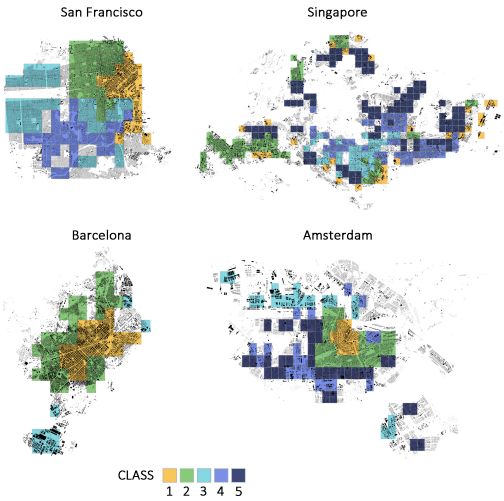
3.3. Data

We use grid cells to partition the four study areas into smaller analysis units. In the initial experiment settings, we use 1km x 1km grids in accordance with the WorldPop dataset (Tatem, 2017). We choose 1km for two reasons: 1) Effective feature extraction — If the grid size is too small, it is difficult to capture continuous figure-ground patterns. Conversely, if the size is too large, the urban patterns are likely to be mixed up and lose important details; 2) Compatibility with the existing data source — Many ground truth data for the application of urban morphology, such as socio-economic activity (Kummu et al., 2018; Lepetit et al., 2023) and air pollution data (Swanson et al., 2022), are in 1km resolution. Therefore, it is crucial to ensure that our method effectively describes urban form at this resolution, facilitating integration with other datasets and supporting broader urban research. Additionally, we conduct an investigation using a 500m x 500m grid size in a subsection to explore benefits and limitations at a finer resolution. A resolution of 500m is also commonly employed in the study of urban morphology (Rode et al., 2014). If a grid has low building coverage, it will be removed from consideration for representativeness.

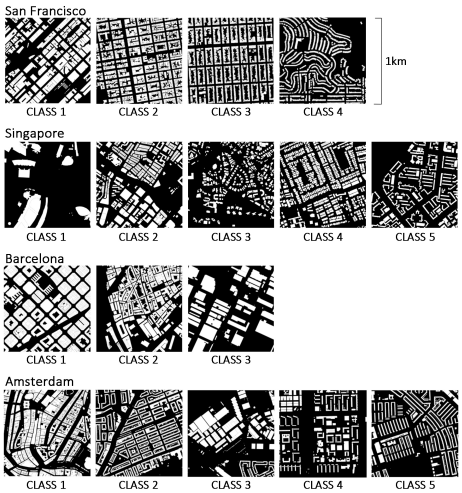
For morphology indicators, the GBMI open dataset provides building indicators such as average footprint areas, length, width, and complexity at the same scale, and it also supplies information on individual building level (Biljecki and

Chow, 2022). The dataset is computed from OSM data, and we use it to complement the visual pattern.

For testing the effectiveness of the proposed method, we label typical morphology classes for the four cities. The labeling follows the historico- geographical approach of morphological region analysis, and is based on multi-layer urban development maps (e.g. land use and building age). Figure 2 shows the figure-ground maps of the identified patterns and their spatial distributions. In total, we labelled 549 patches, with 113 in San Francisco, 261 in Singapore, 108 in Amsterdam, and 67 in Barcelona.



(a) Locations of morphology types.



(b) Morphology types.

Figure 2: San Francisco, CLASS 1: CBD and industrial area, 2: urban block, 3: suburban housing, 4: housing near to nature (mountain, lake). Singapore, CLASS 1: Large public building, 2: CBD and industrial area, 3: landed housing, 4: shophouse, 5: high-rise apartment. Barcelona, CLASS 1: Cerdà's block, 2: organic dense block, 3: industrial area. Amsterdam, CLASS 1: inner city, 2: 19th century expansion, 3: industrial area, 4: housing + public building, 5: modern housing.

3.4. Visual representation learning

3.4.1. SimCLR

We utilise the SimCLR framework developed by Chen et al. (2020) to learn representations (vector embeddings) of morphology patches. It achieves this by maximising the mutual information agreement between different versions of the same image, called augmented views, using a contrastive objective. In simpler terms, it focuses on making similar features within an image appear even more

similar, while at the same time emphasising the differences between features that are not related. It consists of four key components (Figure 3):

1. A pipeline of data augmentation operations, denoted as DA, transforms an original morphology patch p into two augmented views $\tilde{p}_i$ and $\tilde{p}_j$ ($p \sim DA = \{\tilde{p}_i, \tilde{p}_j\}$), which are considered as positive pairs (pairs that should look similar). This process is crucial because it helps the model understand the same image in various ways. In the original design of SimCLR, three consecutive augmentation operations are applied: random cropping and resize, random colour jitter, and random Gaussian blur. In this study, we tailor this data augmentation pipeline (see Section 3.4.3).
2. A convolutional neural network $f(\cdot)$ as an encoder that extracts representations from the two augmented views. This is the part of the model that reads the images, extracts important features, and transforms the views into numerical representation that the model can work with. Here we use ResNet18 (He et al., 2015) that is sufficiently capable of the feature extraction in view of the complexity and image size in our task. We obtain an embedding (feature map) $\mathbf{h_i} = f(\tilde{p}_i) = ResNet18(\tilde{p}_i)$ for each patch. Resnet conducts 3D convolutions for multi-channel imagery, and here we fill the channels with several different indicators to compose such multi-channel images (cf. Section 3.4.2).
3. A projection head $g(\cdot)$ that maps the representations extracted from ResNet18 to a lower-dimension embedding space. To this end, We use a multi-layer perceptron (MLP) with a ReLU activation function σ for nonlinearity, to obtain $\mathbf{z_i} = g(\mathbf{h_i}) = W^{(2)}\sigma(W^{(1)}\mathbf{h_i})$. After this step, 512-dimensional embeddings $\mathbf{h_i}$ are projected to 128-dimensional embeddings $\mathbf{z_i}$. The effectiveness of this projection head has been verified in Chen et al. (2020), i.e. adding such a projection head lifts the meaningfulness of the learned embeddings. After training, we throw away g and use f and representation $\mathbf{h}$ for downstream analyses.
4. A contrastive loss function ℓ is defined to maximise agreement between positive pairs (two augmented views from the same image) and disagreement among negative pairs (two augmented views from different images). This is the rule that guides the learning process, which helps the model figure out how to make similar views look even more similar and dissimilar views look more different. For an example $\tilde{p}_i$, its positive example is another view $\tilde{p}_j$ generated from the same morphology patch, negative examples are the views generated from other patches within the same minibatch $\{\tilde{p}_k\}_{k \neq i,j}$. For

a positive pair of examples (i, j), its loss function is the same as equation (1) in the SimCLR paper (Chen et al., 2020), which also includes the calculation of overall loss value ($\mathcal{L}$) in Algorithm 1. In the end, the model is optimised using this loss function, e.g. with an Adam optimiser.

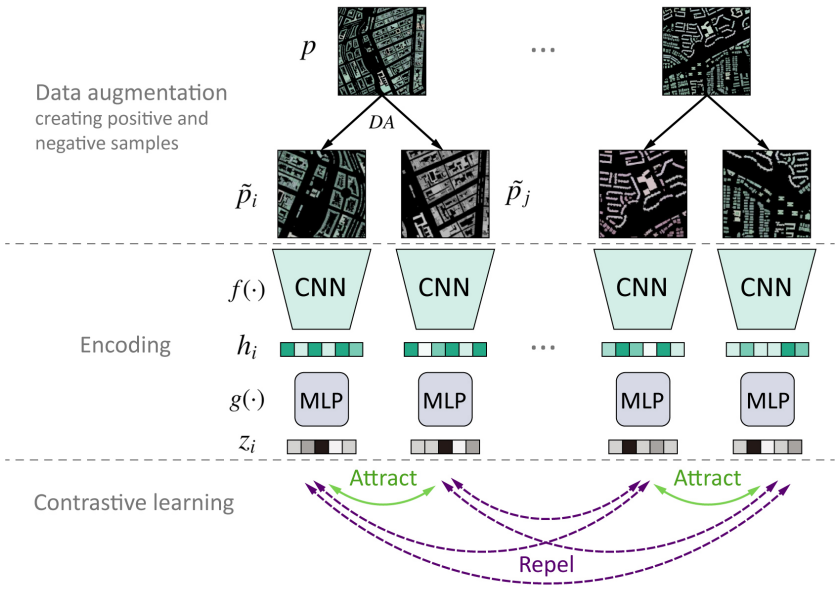


Figure 3: Framework for visual representation learning (SimCLR) applied in discovering urban morphology.

3.4.2. Morphology indicators integration

The geometric properties of individual buildings as well as their spatial layouts are indicative of building morphology. To this end, we enrich the information of morphology patches with three explicit morphometrics, i.e. footprint area, building length, and complexity. In this study, we fill the geometric indicators in the pixels of the morphology patches and compile them through image channels. We name the method as *multi-channel indicators*.

Specifically, we rasterise each patch into a 256 x 256 sized image, and the values from a type of geometric indicator, e.g. footprint area, are filled into the pixel cells that are within the boundaries of building footprints. In this way, each indicator type could form a 256 x 256 image and become a channel. Finally, we stack the three channels pertaining to three indicators to form three-channel

images as the input for CNN. This is akin to the RGB features, while in this case, each channel is an indicator type. Indicator values first undergo a logarithmic transformation and then are re-scaled to [0, 255] (Figure 4).

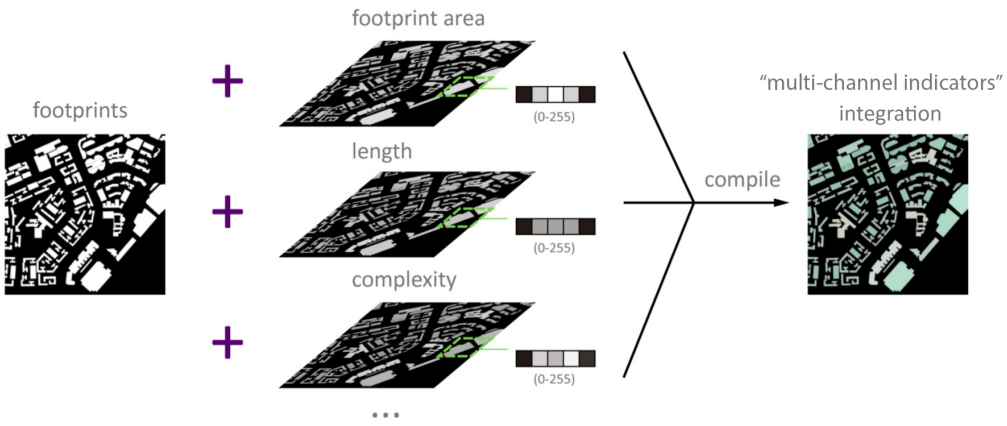


Figure 4: Process of integrating 3 indicators channels with building footprints.

3.4.3. Tailor data augmentation pipeline

The data augmentation step creates two views of each image to act as positive samples (e.g. a *cropped* dog image 'p_i' and a *grayscale* dog image 'p_j' both derived from an original Dog image 'p' are considered similar). Considering the difference between photo recognition (original SimCLR model) and morphology recognition, customising the task-specific data augmentation pipeline (rule for creating positive urban morphology pairs) is pivotal.

In the morphology learning task, different augmentation methods have contextual meaning. *Crop-resize* simulates extracting a small part in the 1 km urban area; *colour transformations* add disturbance to indicators and reduce their importance; *horizontal flip* and *rotation* tell the model to ignore angle differences of buildings; *affine* lets morphological pattern partly fill in the patch, reduces the impact of density differences.

In this study, we devise four varying data argumentation pipelines ($DA1$ – $DA4$), based on the human perception process of urban forms and the empirical results in the study of (Chen et al., 2020). The details of $DA1-DA4$ are manifested in Figure 5.

The first data argumentation pipeline $DA1$ is from the default practice of SimCLR, which has been empirically proved effective for natural images (ImageNet).

The second pipeline *DA*2 removes the colour transformation operators in *DA*1, i.e. *colour jitter* and *grayscale* are removed. The reason of such removal is to keep the geometric indicators embedded in the patch pixels, as colour transformation practically entails the perturbation of geometric indicators.

The third and fourth ones - *DA*3 and *DA*4 - add the *rotate* and *affine* operators to mimic human perception and make the learning process more robust. This is in view of the invariance and robustness of human understanding of different urban forms under rotations and affine transformations of the figure grounds.

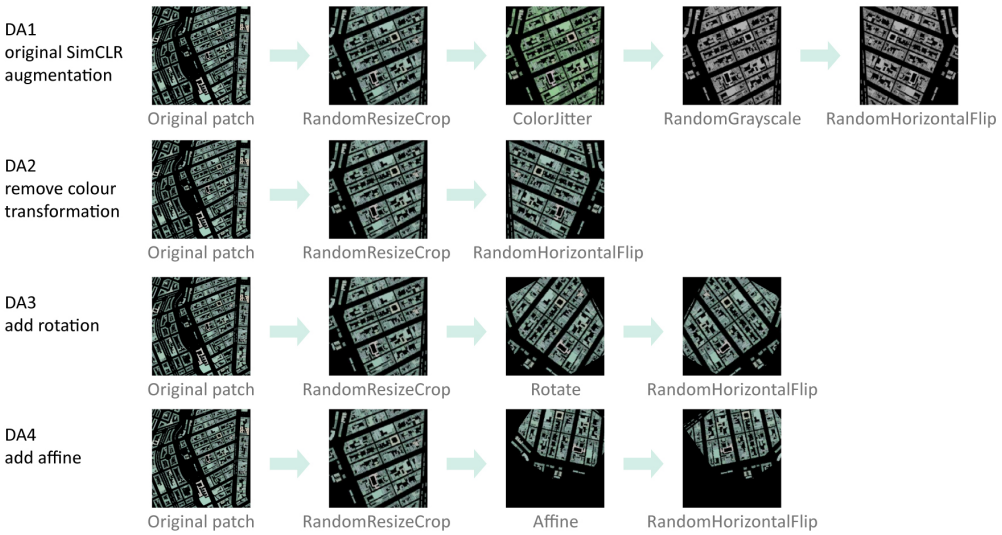


Figure 5: Data augmentation pipelines designed for morphology learning, in a pipeline, each DA step has certain probability to be applied on the original patch.

4. Experiment and results

4.1. Implementation details

We train the model using morphology patches from all cities simultaneously, with a minibatch size N of 128 and a temperature parameter τ of 0.5. The training process spans 200 epochs. Initially, the learning rate is set at 0.01. To optimise training, we use a learning rate scheduler that reduces the learning rate to one-tenth of its current value at the 90th and 110th epochs. Upon completion of the training, we keep the ResNet encoder f and throw away the projection head g. Each patch p then undergoes the same data augmentation process *DA* and the

ResNet encoder f, so as to generate its embedding h used in our downstream analyses.

4.2. Urban morphology classification

Meaningful embeddings should yield higher accuracy in the supervised classification task because they effectively capture the implicit patterns (e.g. configuration and spatial interaction). In order to compare the meaningfulness of the generated embeddings, we test their performances in a supervised urban morphology classification task.

The classification is based on a random forest classifier (100 decision trees). The input for the model is the embeddings of morphology patches; the ground truth data are manually annotated morphology labels (549 labels in total) (See Section 3.3 for detailed information); the outputs are the predicted morphology classes for unknown patches. We split the dataset to training and test sets with the ratio 6:4. The classification task is carried out in the four test cities respectively.

We run the classification model on four embeddings generated by the four data augmentation pipelines ($DA1 - DA4$). These embeddings are based on the *multi-channel indicators* integration method. In addition, for demonstrating the advantage of such method, we create a baseline indicator integration method — *indicator concatenation*. Specifically, we derive several statistics from the three indicators used in the study, forming an indicator feature vector (three-dimensional) for each patch. In the meantime, we rasterise each patch with only building footprint information and feed it into SimCLR, and use $DA1$ pipeline to obtain visual embedding of the patch. Finally, we concatenate the indicator feature vector and visual embedding for each patch as the final representation. This embedding is denoted as $Concat$.

To reduce the randomness of the classification results, we run the classifier for each generated embedding ten times and take the average F1 scores for comparison. Table 1 summarises the embeddings' performances in the four cities in F1 score mean.

Table 1: Summarized results of Random forest classification

	F1 score				
	sf	sg	am	bc	F1-mean
*DA*1	0.84	0.72	0.73	0.74	0.758
*DA*2	0.8	0.82	0.8	0.7	0.780
*DA*3	0.82	0.89	0.8	0.62	**0.783**
*DA*4	0.77	0.82	0.71	0.7	0.750
Concat	0.77	0.73	0.7	0.65	0.713

4.2.1. Best data augmentation pipeline for morphology learning

From the classification task, we observe embedding generated from *DA*1 achieves the best performance in San Francisco and Barcelona. *DA*1 contains steps of colour-jitter and grayscale, meaning if a patch drops the indicator channels it is still considered the same as the original patch. This method is better at distinguishing the visual nuances and is less influenced by morphology indicators.

Such a finding can be explained by the urban form. In San Francisco and Barcelona, buildings are organised in highly regular, gridiron street structures; thus the morphological differences are easily visible via density and patterns. In addition, in the two cities, the individual buildings are largely homogeneous in geometry, making morphometrics less instrumental. Take San Francisco for example, housing in urban blocks, suburbs and housing near to nature (CLASS 1,2,3) are all dominated by single-family homes. In this context, visual features play more important role in differentiating these classes.

The confusion matrix (Figure 6) proves this statement. Compared with the colour-transformation-removed data augmentation (*DA*2), *DA*1 is particularly better at identifying CLASS 2 and 3 in San Francisco.

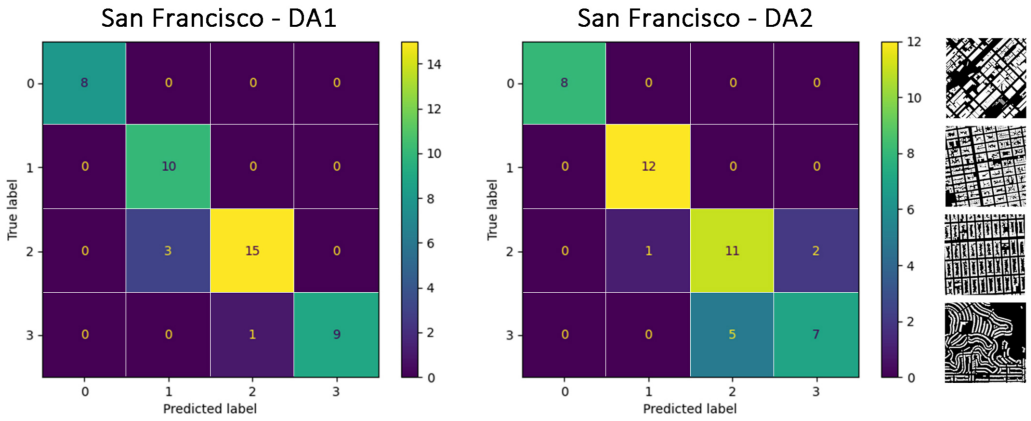


Figure 6: Confusion matrix based on San Francisco DA1 and DA2 embeddings.

However, in cities with less stringent spatial order and more building types such as Amsterdam and Singapore, data augmentation methods that emphasise the importance of indicators are tested more effective. *DA*2 which removes colour transformation upon *DA*1 perform well in Amsterdam, while *DA*3 which adds rotation based on *DA*2 achieve convincing results in both cities.

In Amsterdam, although developments in different historical periods may follow the same fabric as the previous stage, which makes the difference not easily recognisable visually, the building types in each period vary, enabling indicators effective in capturing the characteristics. In Singapore, the urban layout is organised in an organic form, meaning it is different from San Francisco where each morphology type is more uniformed. Within a type in Singapore, the spatial configuration of buildings differ from patch to patch.

The disordered pattern brings challenges to the visual representation learning model, which is predictable — even for humans, finding patterns in the clutter of buildings can be tricky. Therefore we observe that *DA*1 has the lowest performance in Singapore. However, once the indicator channels are preserved while creating positive samples, the performance raises considerably thanks to the city's distinctive building types (Figure 7). Especially in *DA*3, where tolerance in different angles is improved, the F1 score in Singapore raises to 0.89, which is a notable increase.

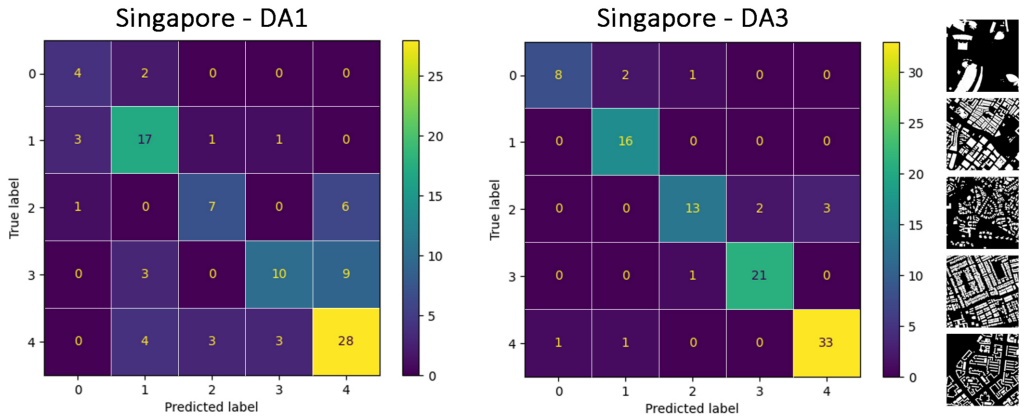


Figure 7: Confusion matrix based on Singapore DA1 and DA3 embeddings.

In order to find the best data augmentation pipeline, we compare the mean and rank of the F1 score (Table 1). *DA*2 and *DA*3 yield similar results in both measurements. Considering *DA*3 performs significantly better in Singapore, a city with highly complex and disordered morphology, it also outperforms *DA*2 in San Francisco, a city with clear spatial logic. In this context, we believe the embedding derived from *DA*3 is the most representative one. Therefore, in the continuation of this study, we opt for *DA*3 in further clustering analysis.

4.2.2. Effectiveness of morphology indicators integration method

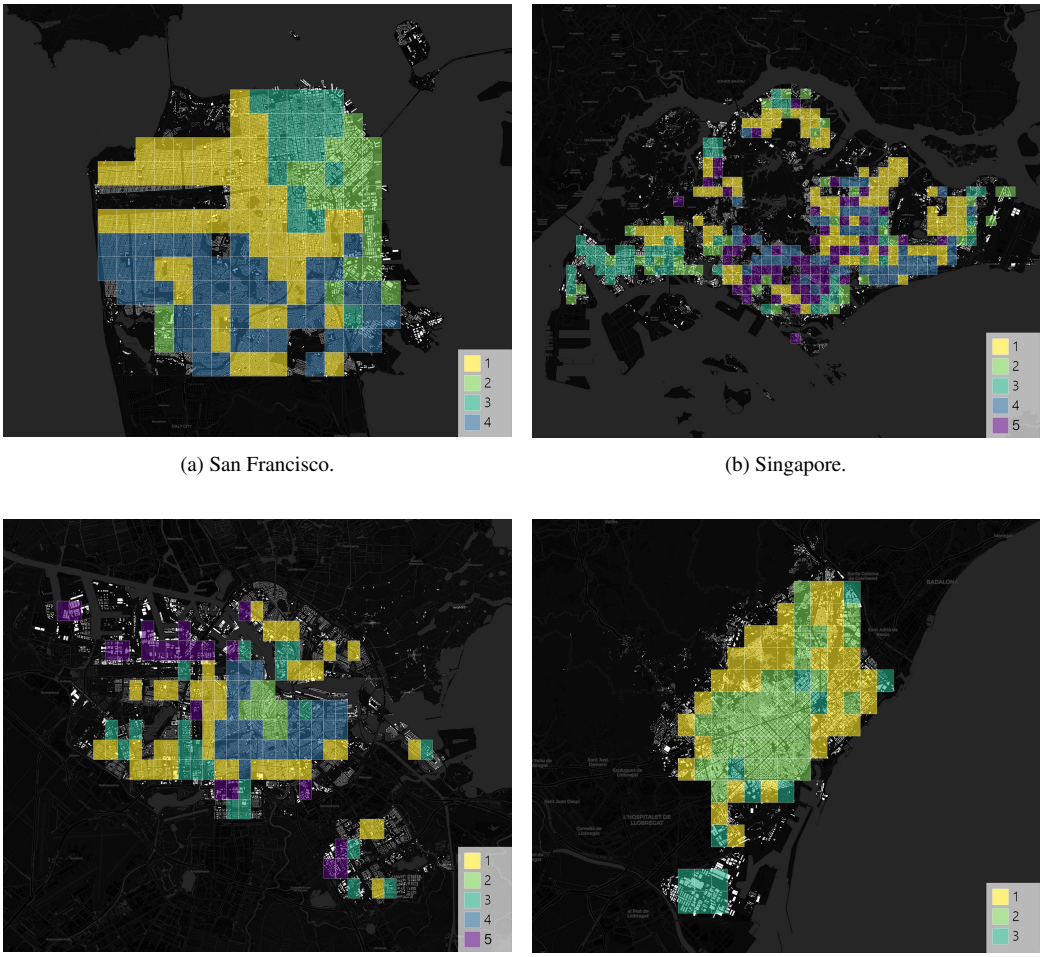
From the summary table (Table 1) we note that in all cities, the *multi-channel indicators* method results in higher F1 scores compared to the simple concatenating indicators method. The *multi-channel indicators* method has high information granularity down to individual building level, and minimises information loss during integration, while preserving both the layout of building groups and geometric information of individual buildings. This novel method is proven valid in the morphology representation learning task.

This method stores morphology information in a flexible yet compact way — the channels can be readily altered, removed, or supplemented. As we identify in the previous section, dropping value disturbance achieves the best result. Such a data augmentation pipeline can be applied to multi-spectrum imagery, i.e. image with more than three channels, which enables further expanding the channel numbers, including more building features such as height, age, colour, and roof shape.

4.3. Discover urban form typologies via clustering

4.3.1. Clustering results

In this section, we test the performance of the learned representation in a popular morphology analysis task – finding urban form typologies. First, we run k-means clustering on the best-performing embedding. To discover localised patterns, we then split the embeddings by city. To determine the best cluster number for each study area, we use the NbClust package in R (Charrad et al., 2014). It computes 30 indices measuring clustering performances (e.g. Silhouette score) and proposes a best clustering scheme. According to the result, the best cluster numbers for San Francisco, Singapore, Amsterdam and Barcelona are 4, 5, 5, and 3, respectively.



(a) San Francisco.

(b) Singapore.

(c) Amsterdam.

(d) Barcelona.

Figure 8: Spatial distribution of morphology types discovered by our approach.

We extract the top 10 nearest patches to each cluster centroid that depict most representative patterns, forming inventories of urban morphology in the four cities (Figure 9).

In our examination of urban morphology, San Francisco (SF) presents four primary types. SF1 embodies the quintessential American 'suburbia', characterised by sparse, single-family homes. SF2, contrasting SF1, represents the bustling city centre, featuring large office buildings and shopping malls arranged in gridiron streets. This type also encompasses industrial zones along the city's east coast. SF3, situated in the transitional zones between urban and suburban areas, presents

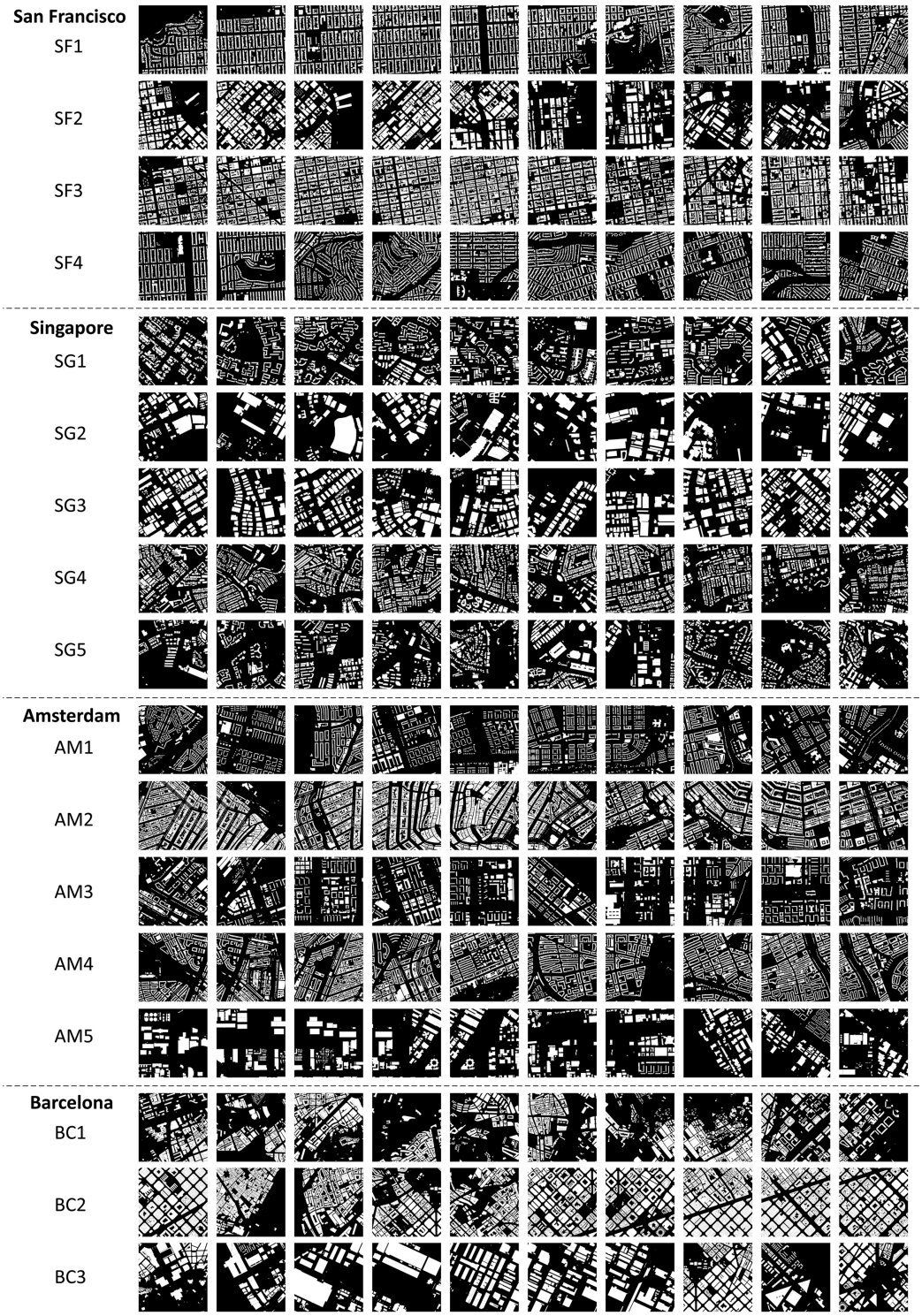


Figure 9: Discovered morphology types in the four cities. The rows, each representing an identified cluster, include 10 examples of patches. For the purpose of clear visualisation, we use the binary figure-ground map for rending.

an intermediate morphological type. Though single-family housing dominates, these residences are denser and often coexist with public facilities and schools. Finally, SF4 bears a striking resemblance to SF1 but displays a greater influence of natural elements, such as rivers, hills, and lakes, that create unique voids in the urban pattern.

Singapore's (SG) morphology offers its own distinct patterns. SG1 exemplifies the 'Towers in the park' design—a concept championed by modernist architects—featuring high-rise residential buildings amid landscaped greenery. SG2 comprises large, publicly accessible buildings and complexes such as Marina Bay Sands and Singapore Changi Airport, often surrounded by wide open spaces. Conversely, SG3, while similar in function to SG2, mainly contains densely placed hotels, offices, and commercial buildings with smaller footprints. Unique to Singapore, SG4 showcases the 'shophouses' concept, compact and narrow structures designed for street-side businesses. SG5 presents a different approach with sparse, organic design emphasising natural elements and including landed houses and detached apartments.

Amsterdam (AM) reflects its historic development through its morphologies. At the city centre, AM2 is dominated by the inner-city canal ring layout, a UNESCO World Heritage Site, featuring slim houses nestled between canals. AM4, representing 19th-century urban expansion, displays buildings arranged in blocks that are more flexible and diverse than SF's grids. AM1 and AM3 depict modern residential forms in the city's west. AM3 also integrates public buildings with larger footprints. Lastly, AM5 represents the city's industrial areas, typified by large, rectangular buildings such as warehouses and factories.

In Barcelona (BC), the influence of Cerdà's square-shaped block plan is evident. BC1 indicates new urban developments on the city's outskirts, mirroring BC2's spatial fabric but featuring longer, multi-story buildings. BC2 epitomises the most iconic urban landscape in Barcelona, blending square blocks with organic town patterns. BC3, clustering in the south, functions as the port and industrial zone. We note that some Cerdà blocks are categorised as BC3, which may be attributed to the coarse modeling approach of OSM building footprint, i.e. a block is represented by a large square instead of a group of buildings. We acknowledge the accuracy of the result is subject to the different mapping conventions and quality of the source building footprint data, which vary around the world (Biljecki et al., 2023). Another issue is the Barrio Gotico area (the old city of Barcelona with narrow medieval streets) is clustered into the same class as the Cerdà blocks. Possible causes for the misclassification are: 1) Small training samples — Barrio Gotico area only accounts for 3 grids; 2) The mix of urban fabrics — Cerdà blocks often

mix with the Barrio Gotico area within 1km grids. We assume a solution to this matter is decreasing the size of 1km grids to 500m, thus increasing the size and quality of training data. We performed the experiment and present the result in the next section.

4.3.2. Effects of grid size and cluster number

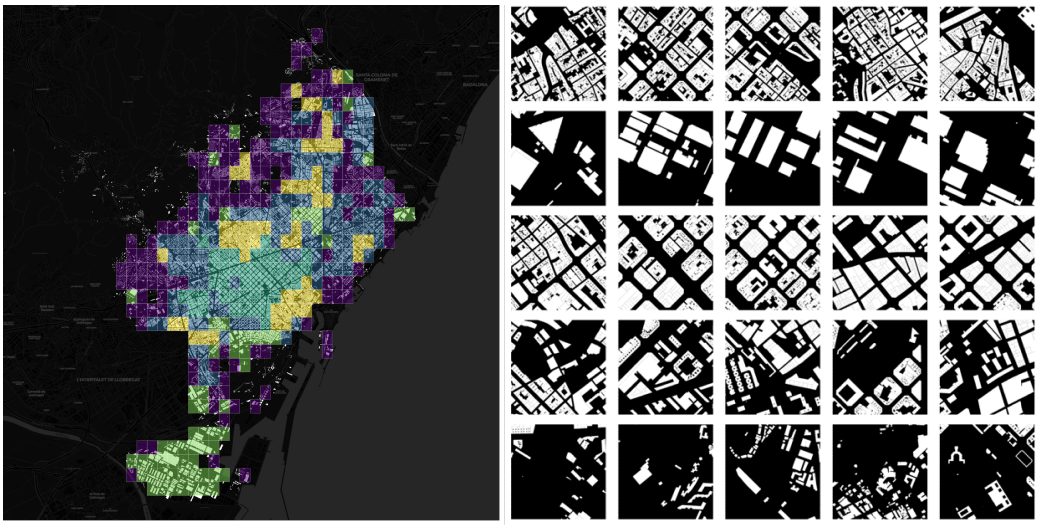
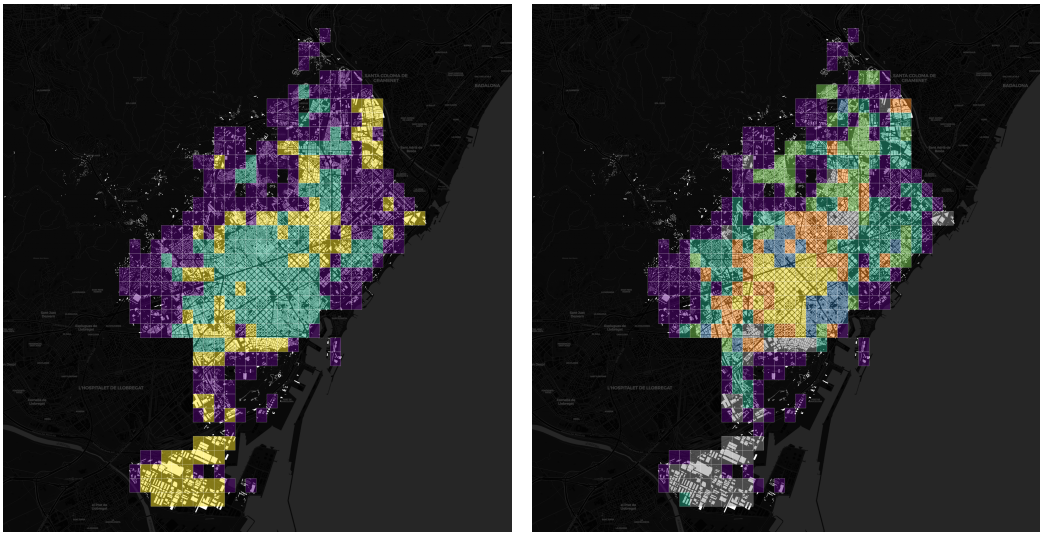


Figure 10: Discovered morphology types based on 500m grids in Barcelona. The rows, each representing a cluster, include 5 examples of patches.

In this section, we reduce the size of the cells to 500m x 500m while keeping the rest of the experimental settings unchanged. The experiment is conducted in Barcelona, and the optimal number of clusters changes from 3 to 5.

Based on Figure 10, we observe a distinction between the Barrio Gotico area and the Cerdà blocks. Furthermore, the urban periphery appears more clearly delineated. Certain densely populated urban neighbourhoods with dense and narrow urban fabric (e.g. the Gràcia area) are differentiated from Cerdà block. These findings validate our assumption that by disentangling mixed urban fabrics and increasing the number of training samples, the clustering results can be improved.



(a) Cluster number:3 (b) Cluster number:7

Figure 11: Discovered morphology types in Barcelona by varying cluster number

Determining the optimal cluster number and selecting the appropriate clustering method are ongoing debates in metrics-based urban typo-morphology analysis. Our morphology representations also encounter the same challenge.

Figure 11 illustrates the results obtained by varying the cluster numbers, with a reduction to 3 clusters on the left and an increase to 7 clusters on the right. When the cluster number decreases, various urban fabrics tend to be grouped together. On the other hand, when the cluster number increases, more intricate urban fabric types are identified. However, it is possible that some of these types may not exhibit significant differences from one another. In addition, these types are not spatially aggregated together due to the absence of spatial proximity enforcement in clustering analysis.

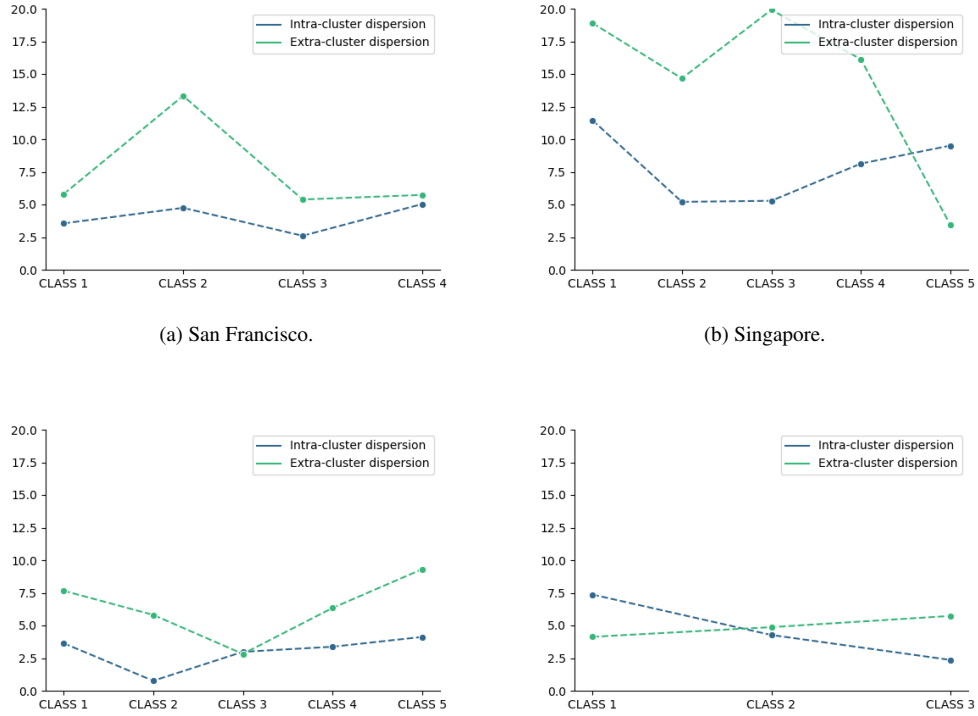
To enhance performance in discovering urban form typologies, employing spatially explicit clustering methods, and using more advanced techniques to determine the best cluster number can be beneficial. From the perspective of representing urban morphology, our proposed method is capable of effectively capturing the variances in figure-ground maps.

4.4. Inner-city homogeneity

In this experiment, we leverage the intrinsic features of clustering results for uncovering the local morphology homogeneity within cities.

We measure homogeneity at two levels. First is the within-cluster level. Homogeneity in this level reveals the degree of regularity and consistency of a morphology type. For example, almost every patch of uniform 'American grids' (SF1, SF3) is identical, therefore they are highly homogeneous, while the patches by 'organic design' (SG1) are dissimilar from each other, so this type is considered more heterogeneous. We use intra-cluster dispersion — the sum of square distances of each point to the corresponding cluster centre (Caliński and Harabasz, 1974). The larger the sum, the more dispersed the cluster, therefore the within-cluster homogeneity is lower.

Second is the cross-cluster level. We measure extra-cluster dispersion by the square distance of each cluster centroid to the centre of all points. If all morphology types within a city are similar, the extra-cluster dispersion values would be low, implying the city has a homogeneous landscape. This index also reflects the uniqueness of a morphology pattern — when a type is further apart from other types, a relatively higher extra-cluster dispersion value would be observed.



(a) San Francisco.

(b) Singapore.

(c) Amsterdam.

(d) Barcelona.

Figure 12: Intra- and extra-cluster dispersion in four cities. Low value suggests more homogeneous urban form.

From Figure 12a, we observe top-down structural planning has shaped an ordered and repetitive urban landscape in San Francisco. The urban street block morphology (SF3) is highly regularised with little variation. This dominating spatial order is only compromised in the face of nature — in the suburb housing type (SF4), the intra-cluster dispersion is slightly higher than in other types. However, from the perspective of urban perception, the CBD and industrial type (SF2) might be most significantly different from the prevailing environment of the city.

The urban environment in Singapore is the opposite of San Francisco, where the level of intra-cluster dispersion varies significantly among morphology types, and each morphology type is strikingly different from another (Fiugre 12b). The residential towers classified under SG1 demonstrate a notable variety in their spatial layouts, followed by sparse housing type (SG5) and shophouses (SG4). The finding is in line with the organic design principle of Singapore, which deliberately creates rich spatial experiences by varying layouts, so as to increase the sense

of space within the land-constrained island.

In Amsterdam, the inner city type (AM2) has the lowest intra-cluster dispersion, implying the underlying strong and consistent spatial logic of the canal ring layout. The industrial type (AM5) is the most dissimilar form to the rest of the city. Interestingly, such a segregation in urban form can be found in almost every type representing industrial sites (SF2, SG3, BC3). This heterogeneity reveals the fact that industry areas are usually enclaves outside of cities and are detached from the local urban fabric.

Similar to San Francisco that is structured by gridiron streets, Barcelona has a homogeneous urban landscape dominated by square blocks (BC2). BC1 patches are usually located near the coast and mountains, and it is hard to find an overarching logic to support its development. Therefore, a higher intra-cluster dispersion is observed.

4.5. Cross-city homogeneity

We further analyse the morphology homogeneity across the four cities. First, we utilise t-SNE (van der Maaten and Hinton, 2008) to cast all the patch embeddings in the four cities to points in a 2D plane for visualisation in Figure 13, so as to obtain intuitions about the cross-city similarities. The closer the points on the 2D space, the more similar they are. Second, we derive all the pairwise cosine distances between the morphology cluster centroids, to concretely understand if a certain morphology pattern in one city occurs in another city. The central point of each cluster is derived by averaging all the embeddings in that cluster.

Observation 1: the influence of globalisation. From our previous analysis, we identified several morphological types associated with business and industrial areas in each city (SF2, SG2 and SG3, AM5, BC3) These types exhibit strikingly similar physical characteristics, as evidenced by their collective positioning in the top-right corner of the t-SNE plane (Figure 13) and their minimal cosine distances from one another (Figure 14). This observation indicates a prevailing similarity across countries in business and industrial landscapes, which could be partially ascribed to the pervasive forces of globalisation. In addition, we observe that these types demonstrate a distinct divergence from the morphology patterns commonly seen in their respective local environments. This exposes a widespread concern: globalisation is likely to bring the risk of dilution of local identity (Savage et al., 2004; Kaymaz, 2013).

Observation 2: identifiable colonial city principles. Historically, the four study areas played adverse roles in colonisation. Amsterdam and Barcelona reflect town planning principles of the colonist, while Singapore and San Francisco,

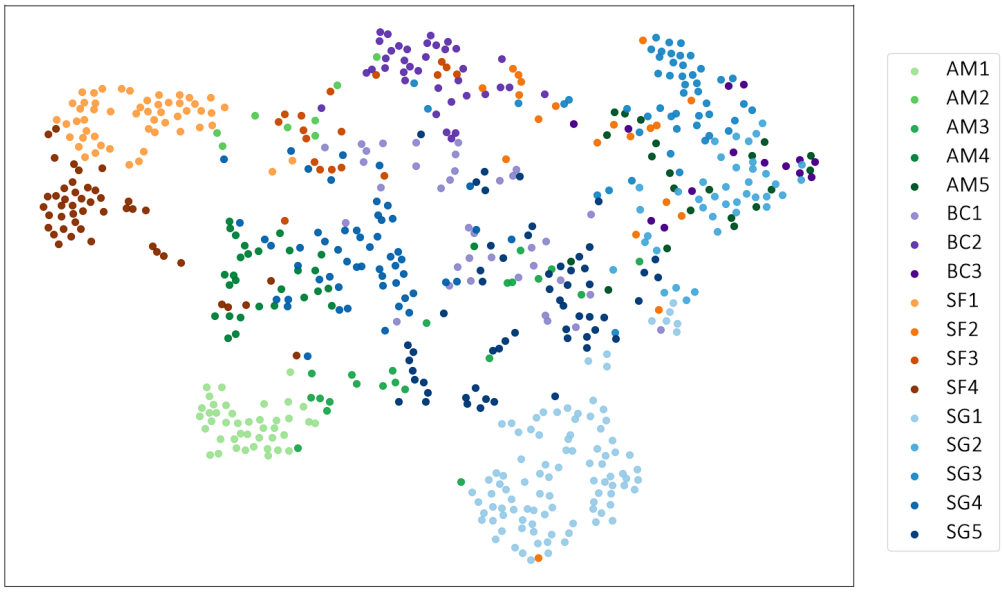


Figure 13: t-SNE visualisation of morphology types. Each point represents a training patch. Similar patches are closer to each other. The color represents the urban form types discovered in the previous section.

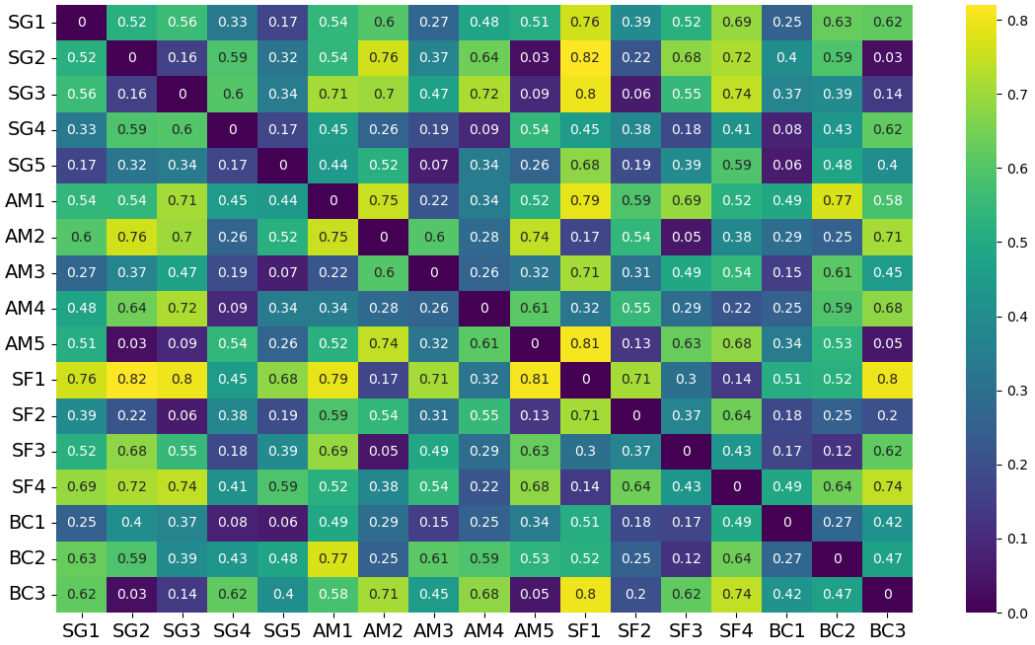


Figure 14: Morphology pairwise cosine distance, the lower the value, the more similar the morphology types.

as former colonies, have built environments that were modified or moulded by the colonisers in the past.There is a discernible consistent morphology *gene* across the four cities. In the t-SNE visualisation (Figure 13), we notice patterns from these cities converging at the middle of the figure. From the colonist cities, traditional European blocks (AM4, BC1) are evident, and from the former colonies, we observe shophouses blending Eastern and Western styles (SG4), alongside American grid patterns that resemble the European blocks (SF3, SF4). The similarities are further validated by the low cosine distance values among the morphology types (Figure 14). The cross-country homogeneity suggests a transfer and adaptation of architectural styles during the colonial period, reflecting a blend of local and foreign influences in urban development.

5. Discussion

5.1. Urban morphology representation learning and limitations

In this section, we discuss several crucial technical points and limitations of our proposed method.

Data augmentation method. Through testing the permutations of data augmentation methods, an optimum pipeline for morphology learning is discovered. It minimises the disturbance to building indicators and increases the tolerance to angle differences, a possible explanation for the good performance is this method is the closest imitation of human cognition in identifying morphology patterns.

Multi-channel indicators. We apply three building indicators in this paper. The method of overlaying indicator channels with building footprint is proved valid with around 20% improvement from simply concatenating the two pieces of information. We speculate that it could be attributed to its ability in reducing information loss, i.e. maintaining the correspondence between geometric information of individual buildings and the spatial layout of building groups.

Since our backbone visual encoder (CNN) does not limit the number of channels, it is possible to incorporate more indicators as additional channels, a common practice in, for example, multi-spectral remote sensing image analyses. However, it is worth noting that using more channels would increase GPU memory consumption and make the training more costly. Additionally, it remains unclear whether adding more similar indicators is beneficial, as it carries the risk of confusing the unsupervised representation learning process.

The size and shape of cells. We compare the clustering performance of using 1km and 500m grids in Barcelona. When using 1km grids, we encounter problems with mixed urban form and small sample sizes. However, when shifting to 500m grids, we notice that those issues are significantly alleviated. Therefore, the optimal grid size depends on the context of the city. For large, rigidly planned cities, using a 1km grid is sufficient to capture the variance of urban form while maintaining computational efficiency. However, for small and hybrid cities, smaller grid sizes like 500m can preserve more details in urban form.

We acknowledge that using grids to partition urban areas is a limitation of this study, because grids do not always align with natural boundaries in cities, and it is common for grids to cut through two different neighbourhoods with varying forms. Another limitation of the current cell shape is that empty areas (could be unbuilt or built up areas with missing data) are not subtracted during encoding. As a result, in downstream typo-morphology clustering analysis, a patch half-filled with a pattern may be recognised differently from a patch fully filled with the same pattern. However, this limitation does not apply when the embeddings are used in some other downstream tasks, e.g. urban function inference, where explicit modelling of empty areas could be beneficial (Li et al., 2023).

Mixed urban forms. Clustering analysis is an important downstream application of our morphology representation, in which we encounter a challenge in dealing with mixed patterns, which could partially be attributed to the grids that we utilise in the study. In addition, urban forms often exhibit gradients rather than falling into simplistic categories, making it difficult for grouping. Our proposed method generates numeric embeddings for each grid cell, resulting in gradual transitions in the embedding space (see Figure 13). To assess if a patch likely represents a mix of class A and B, we can calculate the cosine similarity of each grid cell to the centroids of each morphology type. This approach offers a rough estimation of how similar a patch is to the centroid of class A and B. Nevertheless, it is essential to acknowledge that cosine similarity provides only an approximate indication.

Expert interpretation. Similar to other analyses based on unsupervised learning, our method benefits from expert validation for interpreting morphology discovery results. The influence of several hyper-parameter choices like grid size and the number of clusters on the model's outcomes underscores the necessity of expert input, not only for validating results but also for appropriately selecting these parameters for varied urban contexts.

5.2. Future opportunities

Inspired by the limitations above, we propose several directions for future work.

From 2D morphology to 3D morphology. Our urban environment is not a flat (2D) plane, even though many urban morphology studies have regarded it as such. From the clustering results, we observe that patches in CBD and industrial sites are grouped together. The indicator channels could be enriched with information that better describes the perceived features of the built environment. Buildings raise from the ground shaping undulating skylines, on the facades, varying forms of balconies, decorations styles, and colour paintings give unique identities to the sense of place. Thanks to the flexibility given by information channels, it is possible to add 3D information as multi-channel indicators. Studies have been done for enriching building height data globally (Esch et al., 2022), and computing a rich list of 3D morphology metrics (Labetski et al., 2023). Besides, the building colour tag in OSM allows crowd-sourced gathering of colours. We believe this information is essential for deriving urban morphology types similar to local climate zones (Zhu et al., 2022) but preserving more intricate features of building forms.

Use natural boundaries of urban neighbourhoods. We speculate that blocks partitioned by streets or defined by high-resolution census tracts could possibly be more suitable units for morphology representation learning, which leaves room for future investigation. In this paper, the CNN used takes squares (grid cells) as inputs, and other block shapes would result in gaps within the input. To address the issue of dealing with natural boundaries, a special token is needed to fill the gaps (padding). In addition, using natural boundaries could potentially mitigate the mixed urban form challenge presented in the study.

Consider spatial proximity. The same type of urban form usually spreads across a large urban area, therefore in the 1km x 1km grids, they are likely to appear next to each other. Adding spatial proximity information to representation learning can further reduce errors and outline continuous urban regions with similar morphological features. It could be achieved, among other ways, in the contrastive learning stage — while creating positive samples, neighbouring patches are considered similar to the anchor patch.

Towards integrated, multi-layer urban representation. In this study, we only focus on the building layer, yet the real-world urban environment involves complex interaction of different types of elements including streets, open space, and topography. Compiling the various urban layers to multi-spectrum imagery could be an approach to expanding informativeness of the morphology representations. In addition, other than physical elements, semantic information brought by POIs (Huang et al., 2022) can potentially be incorporated into a multi-modal learning framework, but how to realise it would be an exciting yet challenging mission. We envision this research direction will result in much more informative representations of real-world urban environments, which could be useful in myriads of downstream urban analyses such as housing price prediction, mobility prediction, and population estimation etc.

6. Conclusions

In this study, we introduce a novel method leveraging the state of the art in machine learning to revisit and advance a classic method of urban morphology study — figure-ground analysis. Unlike previous studies using morphometrics or supervised visual classification techniques, we tailor a visual representation learning model to learn latent and meaningful visual urban morphology features in a fully unsupervised manner.

The representation learning process is efficient, scalable, and objective, potentially facilitating global comparative studies. However, our method does not represent a complete divergence from traditional feature engineering method. We build upon morphometrics rooted in previous studies, and integrate them into a novel framework, i.e. through multi-channel indicators. Therefore, the produced visual features entail both the geometric similarity derived from morphometrics, and the spatial configuration of buildings. These two perspectives – the morphological indicators and the spatial configuration – are complementary in providing a more nuanced and comprehensive understanding of urban morphology.

We demonstrate the learned representations are effective in various investigations. Through clustering analysis conducted in four cities, we find that the discovered urban typologies align with urban functions and development history.

Furthermore, we utilise several quantitative measures that assess pattern relationships, evaluating homogeneity at the cluster, city, and cross-city levels. This homogeneity analysis unveils the regularity, diversity, and uniqueness of morphology patterns, shedding light on urban landscape characteristics intertwined with urban design principles.

Expanding our perspective, we identify recurring patterns globally, interpreting them through the lens of global economic and political activities.

In conclusion, we believe this method holds promise as an alternative approach to effectively describe urban morphology patterns.

Acknowledgements

We gratefully acknowledge the valuable comments by the editor and the reviewers. We thank the members of the NUS Urban Analytics Lab for the discussions, and Lubin Bai at Peking University for inputs of the technical aspects. This research was supported by the National University of Singapore under the Start Up Grant R-295-000-171-133 (Large-scale 3D Geospatial Data for Urban Analytics), and the Knut and Alice Wallenberg Foundation (to W.H.).

References

Aibar, E., Bijker, W.E., 1997. Constructing a City: The Cerdà Plan for the Extension of Barcelona. Science, Technology, & Human Values 22, 3–30. doi:10.1177/016224399702200101.

Alexiou, A., Singleton, A., Longley, P.A., 2016. A Classification of Multidimensional Open Data for Urban Morphology. Built Environment 42, 382–395. doi:10.2148/benv.42.3.382.

Bai, L., Huang, W., Zhang, X., Du, S., Cong, G., Wang, H., Liu, B., 2023. Geographic mapping with unsupervised multi-modal representation learning from vhr images and pois. ISPRS Journal of Photogrammetry and Remote Sensing 201, 193–208.

Bansal, P., Quan, S.J., 2022. Relationships between building characteristics, urban form and building energy use in different local climate zone contexts: An empirical study in seoul. Energy and Buildings 272, 112335.

Barke, M., 2018. The importance of urban form as an object of study. Teaching urban morphology , 11–30.

Batty, M., 2009. Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies, in: Meyers, R.A. (Ed.), Encyclopedia of Complexity and Systems Science. Springer, New York, NY, pp. 1041–1071. doi:10.1007/978-0-387-30440-3_69.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35, 1798–1828.

Berghauser Pont, M., Haupt, P., 2005. The Spacemate: Density and the typomorphology of the urban fabric. Nordic Journal of Architectural Research 4, 55–68.

Berghauser Pont, M., Stavroulaki, G., Bobkova, E., Gil, J., Marcus, L., Olsson, J., Sun, K., Serra, M., Hausleitner, B., Dhanani, A., Legeby, A., 2019. The spatial distribution and frequency of street, plot and building types across five European cities. Environment and Planning B: Urban Analytics and City Science 46, 1226–1242. doi:10.1177/2399808319857450.

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

Biljecki, F., Chow, Y.S., Lee, K., 2023. Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. Building and Environment 237, 110295. doi:10.1016/j.buildenv.2023.110295.

Bobkova, E., Berghauser Pont, M., Marcus, L., 2021. Towards analytical typologies of plot systems: Quantitative profile of five European cities. Environment and Planning B: Urban Analytics and City Science 48, 604–620. doi:10.1177/2399808319880902.

Bocher, E., Petit, G., Bernard, J., Palominos, S., 2018. A geoprocessing framework to compute urban indicators: The MApUCE tools chain. Urban Climate 24, 153–174. doi:10.1016/j.uclim.2018.01.008.

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.

Boeing, G., 2021. Spatial information and the legibility of urban form: Big data in urban morphology. International Journal of Information Management 56, 102013. doi:10.1016/j.ijinfomgt.2019.09.009.

Cai, J., Chen, Y., 2022. A novel unsupervised deep learning method for the generalization of urban form. Geo-spatial Information Science 25, 568–587.

Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics 3, 1–27. doi:10.1080/03610927408827101.

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A., 2021. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. arXiv:2006.09882 [cs] arXiv:2006.09882.

Cataldai, G., Maffei, G.L., Vaccaro, P., 2002. Saverio muratori and the italian school of planning typology. Urban morphology 6, 3–14.

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software 61, 1–36. doi:10.18637/jss.v061.i06.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A Simple Framework for Contrastive Learning of Visual Representations. arXiv:2002.05709 [cs, stat] arXiv:2002.05709.

Choi, K., 2018. The influence of the built environment on household vehicle travel by the urban typology in Calgary, Canada. Cities 75, 101–110. doi:10.1016/j.cities.2018.01.006.

Conzen, M.R., 2004. Thinking about urban form: papers on urban morphology, 1932-1998. Peter Lang.

Conzen, M.R.G., 1960. Alnwick, northumberland: a study in town-plan analysis. Transactions and Papers (Institute of British Geographers) , iii–122.

Dibble, J., Prelorendjos, A., Romice, O., Zanella, M., Strano, E., Pagel, M., Porta, S., 2019. On the origin of spaces: Morphometric foundations of urban form evolution. Environment and Planning B: Urban Analytics and City Science 46, 707–730.

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., Zeidler, J., 2022. World Settlement Footprint 3D - A first three-dimensional survey of the global building stock. Remote Sensing of Environment 270, 112877. doi:10.1016/j.rse.2021.112877.

Fleischmann, M., Arribas-Bel, D., 2022. Geographical characterisation of british urban form and function using the spatial signatures framework. Scientific Data 9, 546.

Fleischmann, M., Feliciotti, A., Romice, O., Porta, S., 2022. Methodological foundation of a numerical taxonomy of urban form. Environment and Planning B: Urban Analytics and City Science 49, 1283–1299.

Gil, J., Beirão, J., Montenegro, N., Duarte, J., 2012. On the discovery of urban typologies: Data mining the many dimensions of urban form. Urban Morphology 16, 27–40.

Godfrey, B.J., 1997. Urban Development and Redevelopment in San Francisco*. Geographical Review 87, 309–333. doi:10.1111/j.1931-0846.1997.tb00077.x.

Hamilton, W.L., Ying, R., Leskovec, J., 2018. Representation Learning on Graphs: Methods and Applications. arXiv:1709.05584 [cs] arXiv:1709.05584.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum Contrast for Unsupervised Visual Representation Learning. arXiv:1911.05722 [cs] arXiv:1911.05722.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs] arXiv:1512.03385.

Hebbert, M., 2016. Figure-ground: History and practice of a planning technique. Town Planning Review 87, 705–728.

Hecht, R., Herold, H., Meinel, G., Buchroithner, M., 2013. Automatic derivation of urban structure types from topographic maps by means of image analysis and machine learning, in: 26th international cartographic conference, pp. 1–18.

Huang, W., Cui, L., Chen, M., Zhang, D., Yao, Y., 2022. Estimating urban functional distributions with semantics preserved poi embedding. International Journal of Geographical Information Science , 1–26.

Ignatieva, M., Stewart, G.H., 2009. Homogeneity of urban biotopes and similarity of landscape design language in former colonial cities, Cambridge University Press.

Jacobs, A.B., 1993. Great Streets. Technical Report qt3t62h1fv. University of California Transportation Center.

Jochem, W.C., Leasure, D.R., Pannell, O., Chamberlain, H.R., Jones, P., Tatem, A.J., 2021. Classifying settlement types from multi-scale spatial patterns of building footprints. Environment and Planning B: Urban Analytics and City Science 48, 1161–1179.

Kaymaz, I., 2013. Urban landscapes and identity, in: Advances in landscape architecture. IntechOpen.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems, Curran Associates, Inc.

Kropf, K., 2018. The Handbook of Urban Morphology. John Wiley & Sons.

Kummu, M., Taka, M., Guillaume, J.H., 2018. Gridded global datasets for gross domestic product and human development index over 1990–2015. Scientific data 5, 1–15.

Labetski, A., Vitalis, S., Biljecki, F., Arroyo Ohori, K., Stoter, J., 2023. 3D Building Metrics for Urban Morphology. International Journal of Geographical Information Science 37, 36–67.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation 1, 541–551. doi:10.1162/neco.1989.1.4.541.

Lepetit, Q., Viguié, V., Liotta, C., 2023. A gridded dataset on densities, real estate prices, transport, and land use inside 192 worldwide urban areas. Data in Brief 47, 108962.

Li, N., Quan, S.J., 2023. Identifying urban form typologies in seoul using a new gaussian mixture model-based clustering framework. Environment and Planning B: Urban Analytics and City Science , 23998083231151688.

Li, Y., Huang, W., Cong, G., Wang, H., Wang, Z., 2023. Urban region representation learning with openstreetmap building footprints, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1363–1373.

Liu, T.K., 2015. Planning & Urbanisation in Singapore: A 50-Year Journey, in: 50 Years of Urban Planning in Singapore. WORLD SCIENTIFIC. World Scientific

Series on Singapore's 50 Years of Nation-Building, pp. 23–44. doi:10.1142/9789814656474_0002.

Liu, Y., Zhao, K., Cong, G., 2018. Efficient Similar Region Search with Deep Metric Learning, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, New York, NY, USA. pp. 1850–1859. doi:10.1145/3219819.3220031.

van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. Journal of Machine Learning Research 9, 2579–2605.

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] arXiv:1709.02939.

Moudon, A., 1997. Urban morphology as an emerging interdisciplinary field. Urban Morphology 1, 3–10.

Nasar, J.L., 1989. Perception, Cognition, and Evaluation of Urban Places, in: Altman, I., Zube, E.H. (Eds.), Public Places and Spaces. Springer US, Boston, MA. Human Behavior and Environment, pp. 31–56. doi:10.1007/978-1-4684-5601-1_3.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987. doi:10.1109/TPAMI.2002.1017623.

Oliveira, V., Yaygın, M.A., 2020. The concept of the morphological region: developments and prospects .

Perez, J., Fusco, G., Araldi, A., Fuse, T., 2020. Identifying building typologies and their spatial patterns in the metropolitan areas of Marseille and Osaka. Asia-Pacific Journal of Regional Science 4, 193–217. doi:10.1007/s41685-019-00127-6.

Quan, S.J., Li, C., 2021. Urban form and building energy use: A systematic review of measures, mechanisms, and methodologies. Renewable and Sustainable Energy Reviews 139, 110662.

Reps, J.W., 1965. The Making of Urban America: A History of City Planning in the United States. Princeton University Press.

Rode, P., Keim, C., Robazza, G., Viejo, P., Schofield, J., 2014. Cities and energy: urban morphology and residential heat-energy demand. Environment and Planning B: Planning and Design 41, 138–162.

Rowe, C., Koetter, F., 1984. Collage City. MIT Press.

Salazar Miranda, A., 2020. The shape of segregation: The role of urban form in immigrant assimilation. Cities 106, 102852. doi:10.1016/j.cities.2020.102852.

Savage, M., Bagnall, G., Longhurst, B.J., 2004. Globalization and belonging. Sage.

Savini, F., Boterman, W.R., van Gent, W.P.C., Majoor, S., 2016. Amsterdam in the 21st century: Geography, housing, spatial development and politics. Cities 52, 103–113. doi:10.1016/j.cities.2015.11.017.

Schirmer, P.M., Axhausen, K.W., 2019. A Multiscale Clustering of the Urban Morphology for Use in Quantitative Models, in: D'Acci, L. (Ed.), The Mathematics of Urban Morphology. Springer International Publishing, Cham. Modeling and Simulation in Science, Engineering and Technology, pp. 355–382. doi:10.1007/978-3-030-12381-9_16.

Serra, M., Psarra, S., O'Brien, J., et al., 2018. Social and physical characterization of urban contexts: Techniques and methods for quantification, classification and purposive sampling. Urban Planning 3, 58–74.

Song, Y., Knaap, G.J., 2007. Quantitative classification of neighbourhoods: The neighbourhoods of new single-family homes in the portland metropolitan area. Journal of urban design 12, 1–24.

Stojnić, V., Risojević, V., 2021. Self-Supervised Learning of Remote Sensing Scene Representations Using Contrastive Multiview Coding. arXiv:2104.07070 [cs] arXiv:2104.07070.

Swanson, A., Holden, Z.A., Graham, J., Warren, D.A., Noonan, C., Landguth, E., 2022. Daily 1 km terrain resolving maps of surface fine particulate matter for the western united states 2003–2021. Scientific Data 9, 466.

Tatem, A.J., 2017. WorldPop, open data for spatial demography. Scientific Data 4, 170004. doi:10.1038/sdata.2017.4.

Trancik, R., 1991. Finding Lost Space: Theories of Urban Design. John Wiley & Sons.

van Strien, M.J., Adrienne Grêt-Regamey, 2022. Unsupervised deep learning of landscape typologies from remote sensing images and other continuous spatial data. Environmental Modelling & Software 155, 105462. doi:10.1016/j.envsoft.2022.105462.

Vanderhaegen, S., Canters, F., 2017. Mapping urban form and function at city block level using spatial metrics. Landscape and Urban Planning 167, 399–409. doi:10.1016/j.landurbplan.2017.05.023.

Wang, J., Biljecki, F., 2022. Unsupervised machine learning in urban studies: A systematic review of applications. Cities 129, 103925. doi:10.1016/j.cities.2022.103925.

Wheeler, S.M., 2015. Built Landscapes of Metropolitan Regions: An International Typology. Journal of the American Planning Association 81, 167–190. doi:10.1080/01944363.2015.1081567.

Whitehand, J.W.R., Gu, K., Whitehand, S.M., Zhang, J., 2011. Urban morphology and conservation in China. Cities 28, 171–185. doi:10.1016/j.cities.2010.12.001.

Wolf, L.J., Knaap, E., Rey, S., 2021. Geosilhouettes: Geographical measures of cluster fit. Environment and Planning B: Urban Analytics and City Science 48, 521–539.

Xia, C., Zhang, A., Yeh, A.G., 2022. The varying relationships between multidimensional urban form and urban vitality in chinese megacities: Insights from a comparative analysis. Annals of the American Association of Geographers 112, 141–166.

Yap, W., Janssen, P., Biljecki, F., 2022. Free and open source urbanism: Software for urban planning practice. Computers, Environment and Urban Planning 96, 101825. doi:10.1016/j.compenvurbsys.2022.101825.

Ye, Y., Li, D., Liu, X., 2018. How block density and typology affect urban vitality: An exploratory analysis in Shenzhen, China. Urban Geography 39, 631–652. doi:10.1080/02723638.2017.1381536.

Zhang, P., Ghosh, D., Park, S., 2023. Spatial measures and methods in sustainable urban morphology: A systematic review. Landscape and Urban Planning 237, 104776.

Zhu, Q., Liao, C., Hu, H., Mei, X., Li, H., 2020. Map-net: Multiple attending path neural network for building footprint extraction from remote sensed imagery. IEEE Transactions on Geoscience and Remote Sensing 59, 6169–6181.

Zhu, X.X., Qiu, C., Hu, J., Shi, Y., Wang, Y., Schmitt, M., Taubenböck, H., 2022. The urban morphology on our planet–global perspectives from space. Remote Sensing of Environment 269, 112794.