Towards a new image archive for the built environment

Kartikeya Date and Yael Allweil
Technion, Israel Institute of Technology, Israel

Abstract
The ever-growing online corpus of images of the built environment, on social media and mapping platforms, offers a new kind of archive of the built environment. Recent advances in computer vision, specifically convolutional neural networks, offer new ways of querying and analyzing large image corpuses. In this paper, we propose a new method by which historians of the built environment can use these vast image corpuses in their study, enabling new research questions. To demonstrate proof of need, we report on an ongoing case study in Tel Aviv that attempts to show the feasibility of our proposed method for enabling a Historic Urban Landscapes (HUL)-based approach to the study of the built environment. In so doing, we show how such image corpuses could potentially form a new type of archive for architectural and urban history.

Keywords
Archives, built environment, neural networks, architectural history, methods

Introduction: The problem of changing datasets
For scholarly inquiry, historians of architecture and urbanism depend on archives of drawings, historical photographs and other visual data. Urban geographers and urban morphologists also depend on access to large urban datasets and data repositories of visual sources. Electronic modes enabling users to produce, store, and retrieve photographs offer a significant shift in the structure and substance of the pictorial archive, with potentialities for posing new research questions.

In 2015, the New York Times reported that by 2017, about 1.3 trillion digital photographs would be taken annually all over the world, of which 80% would be taken on smartphones (Heyman, 2015). This vast, growing trove of pictures offers an unprecedented record of the contemporary world. Systems like Google StreetView provide a more systematic photographic record of the built environment using multi-perspective panoramas (Roman et al., 2004). This new pictorial record offers new opportunities for scholars of the built environment.
environment: architects, historians, planners, social scientists. They serve as “the digital skin of cities” (Rabari and Storper, 2015).

Unlike text-based data repositories, for which both archival and retrieval methods are well developed, archives of visual data do not yet provide methods for content-based semantic collection and search. This pictorial content is most commonly organized using tags; textual keywords describing the content of a photograph, used both in historical archives and in image collections like ImageNet or ArtStor. The absence of universally agreed keywords means that images are tagged by users according to their own purposes. Tagging is thus semantically uncertain and irregular, limiting the usefulness of this approach. At the same time, the distributed production of text and images by users of social media platforms has produced vast, hitherto unavailable corpuses. They have “democratized” the production of these corpuses (Gane and Beer, 2008; Harris, 2014; Manoff, 2004).

Archives of the built environment consist of collections of architectural documents. These tend to be available primarily for important buildings that have shaped the architectural canon. This approach to what is worthy of archiving characterizes scholarship in architecture history and urban planning (Fletcher, 1931; Frampton, 1981; Jencks, 2002). Over the past decade, the definition of built heritage has evolved from an object-based approach towards a more inclusive urban landscape approach, which includes expanding inquiry of cultural significance beyond iconic buildings and historical fabrics to include large-scale assessment of the built environment (UNESCO, 2012; van Oers, 2010; Veldpaus et al., 2013). A Historic Urban Landscape (HUL) approach requires new types of archives and new data collection approaches.

These new methods must involve novel ways to query image corpuses. They must also involve new conceptual mechanisms to understand how image corpuses represent the built environment.

Recent advances in Computer Vision, specifically Content-Based Image Retrieval (CBIR) and, more recently, Convolutional Neural Networks (CNN) have made feasible the semantic analysis of large picture corpuses (Datta et al., 2008; Russakovsky et al., 2015). They promise the use of large image corpuses of the built environment as a new kind of archive for the study of the built environment. In this paper, we propose a method that adopts these advances on a large image corpus of building facades, to enable the study of the urban environment in ways that hitherto have not been possible. We propose that such a method locates these large image corpuses at the basis of a new kind of urban archive.

**Literature review**

*Methodological advances and changing data repositories*

Urban geographers have adopted online image corpuses more readily than architectural or urban historians. Image or text corpuses, produced across social media, prompted new digital–visual methods and artifacts of inquiry among geographers and allied scholars of the built environment (Leszczynski, 2019). For instance, Rose and Willis studied Twitter accounts associated with institutions and programs promoting the idea of “smart cities” to learn how the idea of “smart” was constituted in these hyper-seen tweeted images (Rose and Willis, 2019). They used Image Plot, which has the ability to search image (and tweet) metadata and image features like saturation, hue or brightness found in CBIR approaches (these are discussed later in this literature review). Hochman and Manovich used Instagram to study spatial-temporal “visual signatures” of 13 cities. In their extended examination of the patterns of Tel Aviv using Image Plot, they show how different parts of the city assume
importance or fade away at the different temporal scales, from the time of the day to the day of the year (Hochman and Manovich, 2013). Boy and Uitermark created a dataset of 400,000 Instagram posts located in Amsterdam to study how the city is represented by its people on Instagram (Boy and Uitermark, 2017). Each of these studies uses geotagged and timestamped corpuses produced on social media platforms to study and visualize spatio-temporal patterns using programs like Image Plot. The attention given to new digital–visual methods in adjacent visual fields like geography (Leszczynski, 2019), and studies that use archives of tweets and Instagram posts, indicate the potential of such new methods for architectural and urban history.

Scholarship in urban humanities has been concerned with developing methods that use computational or digital structured records, like geographic information systems databases and mapping applications. These inquiries at the intersection of history and geographic information systems have produced novel representations of urban history such as Deep Maps (Bodenhamer, 2013; Bodenhamer et al., 2015).

The data made available by the archive train the scholar on what questions can be asked and shape the possibilities of inquiry and interpretation (Farge and Davis, 2013). With the democratization of corpuses through online social media and mapping platforms, the ways in which new data repositories shape the possibilities of inquiry depend on the methodologies and analytical capabilities we deploy to explore and study these vast corpuses of “big data.” The rapidly expanding size of electronic archives remains a significant practical and conceptual problem (Lavee, 2017, 2019; Milligan, 2016). The method we propose requires semantic understanding of images as representations of the built environment.¹

**Representations of the built environment**

Representational systems lie at the heart of almost all systematic inquiry and help develop abstractions. Abstraction and ambiguity in representation systems enable emergent properties and alternative interpretations to be identified. This basic conception of a representation can be found in different disciplines such as engineering (Bucciarelli, 1988; Ferguson, 1994), scientific practice (Daston and Galison, 1992; Latour, 1986; Lynch and Woolgar, 1990), design ideation (Goel, 1995; Goldschmidt, 1991), mapping and geography (Harley, 2001; Pickles, 1995). A representation denotes an object but does not necessarily visually resemble the object (Goodman, 1976). It involves a dialectical relationship between its appearance and the content it is supposed to convey, including two main features: a structure, and a relationship to the object it denotes (Bernheimer, 1961; Wollheim, 1977).

Architectural practice and inquiry into the built environment have relied on representations of the built environment for at least five centuries (Kalay, 2004; Kostof, 2000). Historians have shown how representational systems and technologies have shaped the production and dissemination of knowledge about the built environment (Carpo, 2001; Perez-Gomez and Pelletier, 2000). In his canonical work on the legibility of urban environments, Kevin Lynch argued that people develop mental images that depend on five elements—landmarks, nodes, paths, districts, and edges (Lynch, 1960). These elements were essentially geometrical. Subsequent work identified other types of elements—symbolic, cultural, personal—which also enable legibility or help describe forms which constitute the structure of the environment (Golledge and Spector, 1978; Habraken and Teicher, 2000). Haken and Portugali proposed these non-geometrical elements be called semantic urban elements (Haken and Portugali, 2003). Historians use systems of elements to develop taxonomies and hierarchies which make sense of the built environment, ranging from geometrically based classical orders to more complex conceptualizations which include form, use,
and/or symbolism of various types (Jencks, 2002; Pevsner, 1976; Wittkower, 1952). Christopher Alexander’s work on patterns marks a significant milestone in the scholarship from a standpoint of architectural practice. Alexander’s later emphasis on appropriateness demonstrates the difficulties inherent in developing and justifying such typological hierarchies (Alexander, 1977; Protzen and Alexander, 1980).

The multidisciplinary field of urban morphology, which has traditionally been concerned with the study of long-term transformation of urban environments by considering the hierarchy of relations structured around fundamental physical elements such as streets or plots (Oliveira, 2016), centrally concerned with representational questions and has been influenced by increasingly ubiquitous large-scale datasets and the new types of modeling possibilities they give rise to (D’Acci, 2019). As Behnisch, Hecht, and Herold have observed, these new datasets set up the potential for developing new metrics to study urban forms and the ways in which they change (Behnisch et al., 2019). Quantitative approaches to classifying urban forms and identifying building typologies in cities have typically involved the use of GIS datasets and have been defined in terms of geometrical relationships like scale, distance, convexity or elongation (Berghauser Pont et al., 2019; Colaninno et al., 2011; Perez et al., 2018). Recently, Araldi and Fusco proposed a method for studying the urban fabric quantitatively using data from the pedestrian standpoint (Araldi and Fusco, 2017). All these approaches use geometric relationships to mediate between conceptual feature sets and built environment datasets (typically in the form of GIS layers).

### Classifying images

Images are arrays of two-dimensional pixels. Traditionally, image classification has been done using textual tags. Labeled image corpuses can be queried using these tags. Tags describe some conceptual or semantic information about the image and can be part of a system of tags by which tags are related in known ways. This system has two main limitations. First, if a better system of tags is subsequently conceived, the existing tagging system must be discarded, and the process must begin anew. Second, tagging some images in a corpus provides no insight into other untagged images in it. A textual tag does not, by itself, enable an inquiry into the corpus.

Techniques in Computer Vision have enabled more advanced approaches to classifying images. These approaches use features of the image itself to classify the image. The broad goal of Computer Vision is “to use the observed image data to infer something about the world” (Prince, 2012: 55). CBIR techniques involve interrogating digital picture archives by their visual content based on features that define a visual property of a picture. Color, texture, and shape are the most common examples of features in CBIR. Each pixel consists of its location and its red, green, and blue (RGB) values. The pixels which make up a picture or some part of it can be summarized based on color. Texture involves repetitive patterns of surfaces within the image. Shape features are used especially for segments of images and involve the use of edge-detection methods. These features can be compared using methods like clustering (k-means, hierarchical and others) or classified using methods such as Bayesian classification, support vector machines, k-nearest neighbors, quadtrees (Datta et al., 2008).

The Streetscore project, which uses Google StreetView data to predict the safety of a built environment based on what it looks like to a human, uses color and texture histograms among a host of other features (Naik et al., 2014, 2017). This approach was also used to identify properties of pictures of the built environment, such as memorability. It was found that the memorability of a picture of the built environment could be predicted based on features found in the pictures (Isola et al., 2014).
Doersch et al. sought to identify building elements—windows, balconies, street signs—distinctive to a given city or geospatial area, given a large corpus of geotagged images of buildings. They used Google StreetView images from 12 cities (10,000 pictures from each). Each picture was broken down into non-overlapping patches. Frequently occurring patches in a given city which do not appear in other cities were identified using nearest-neighbor classification. This method was described as an approach to “computational geo-cultural modeling” (Doersch et al., 2015).

CNN offer a different approach to image classification. A neural network is a universal function approximator. Instead of using a known set of rules on a given input to produce an output, a neural network takes a set of known mapped inputs and outputs to approximate a mathematical model which satisfies this mapping (Goodfellow et al., 2016). This model can then be used to estimate outputs for unknown inputs similar to inputs in the known set. Unlike CBIR-based approaches, in CNN, features of interest do not have to be explicitly modeled in terms of some geometric property of the image.

CNNs were first used successfully to recognize handwritten zip code digits and then handwritten characters (LeCun et al., 1989, 1998). Within the last decade, the ImageNet database of more than 14 million pictures annotated to describe more than 20,000 categories has been developed as a benchmark problem for image recognition and classification algorithms (Deng et al., 2009). Until 2012, the best performing algorithms on this benchmark problem had an error rate of 25%. In 2012, Krizhevsky and colleagues improved this error rate to 15% using a deep CNN and graphics processing units (Krizhevsky et al., 2012). Since then CNN models have further reduced this error rate down to 2.3% (Chollet, 2017; He et al., 2016; Russakovsky et al., 2015; Simonyan and Zisserman, 2015; Szegedy et al., 2015). A variant of the basic image classification problem is the object detection problem in which the task is to locate an object within an image.

A second class of CNN models known as fully connected networks have been developed for the purpose of enabling semantic segmentation of an image. Unlike image classification, the purpose of semantic segmentation is to categorize every pixel in an image as belonging to one specific semantic class. For example, each pixel in a streetscape would be categorized depending on what it depicts—a person, a building, a road, a tree, the sky or a vehicle (Long et al., 2015).

The current state of the art of the CNNs for image classification or segmentation, as well as object detection, suggests that CNNs are uniquely well equipped to the problem of interrogating an image corpus using a set of semantic features. CNN has already begun to be used to answer questions about the built environment. They were used with relative success for land-use identification tasks using a combination of satellite images and Google StreetView data (Kang et al., 2018). Recently, the URBAN-i model has been developed to identify informality and slums in urban scenes (Ibrahim et al., 2019).

Proposed methodology

Purpose

We propose a method that uses CNN as the mechanism for using semantic feature sets (called “semantic swatches”) for interrogating large image corpuses of the built environment. The traditional approach in architectural history for interrogating such large corpuses has been to use a system of textual tags or labels. As discussed previously in this paper, this approach is limited because (1) textual labels are external interpretations rather than semantic readings of the recorded object, (2) labeling some images provides no insight into other
images in the corpus, and (3) this limits the scholar’s ability to update a swatch based on an inquiry into the image corpus.

The idea of updating a swatch is significant. The semantic feature sets used by a historian or urban scholar represent ideas and concepts of interest to the inquiry and an understanding of the subject matter. The ability to test these features and then update them is central to advancing any inquiry. As seen in the previous section, this inquiry has hitherto been done either qualitatively, such as in the work of Lynch, Pevsner or Jencks for their respective studies, or quantitatively using geometric models (either CAD or GIS in studies by urban morphologists). The purpose of the current method is to develop a way to test features and update them—to advance the understanding of semantic feature sets—by directly interrogating image corpuses using CNN.

Outline of the method

Consider a hypothetical study of residential building facades in a city. The study aims at contributing towards a historical account of the city’s urban fabric of residential buildings, or at producing a systematic account of architectural-urban processes such as building additions by specific by-laws. These residential buildings constitute the majority of urban streetscape and are accessible in image form using a service like Google StreetView. If such a dataset were to consider all the residential neighborhoods of the city, it would involve hundreds, if not thousands, of facades.

We choose to approach this large corpus of images using a preliminary set of expert-identified semantic swatches. This idea of using a feature set to interrogate the built environment is not new. In previous attempts to survey architectural history, a common approach is to first identify the organizing features of such a survey (Fletcher, 1931; Kostof, 1985; Morris, 1979; Pevsner, 1976). As the survey progresses, the organizing features are refined and updated. The final work, when it is presented, reports the most advanced version of the feature set.

Doersch et al.’s study of Paris and 11 other cities (Doersch et al., 2015) involved trying to identify windows and balconies distinctive to Paris within a large corpus of geotagged images. The (successful) hypothesis here was that there are windows or balconies which are distinctive to the city of Paris—thus uniquely identify a building being in Paris.

The building facades in this hypothetical corpus have different types of windows, balconies, cladding. They also have different types of ornamentation, structural elements, non-structural elements. Some might have rooftop penthouses; others might have stilts at the ground level. Architecturally relevant features of a building facade exist at different levels of detail, from the individual window, to the whole facade. Some of these features are determined by physical form, while others are not immediately apparent based purely on the physical appearance. These features represent different historical periods in the city’s urban-economic development, and urban-architectural culture.

Architecture and urban historians who seek to survey and analyze the residential neighborhoods of a city using large image corpuses—rather than the traditional approach of generalizing from select iconic exemplars (Jencks, 2002)—require a way of categorizing and organizing these facades based on a hierarchical set of features or feature groups. For example, certain arrangements of windows and balconies indicate a particular period in the city’s urban development, as can be seen in Paris’ Boulevard-facing glazed French windows of the 1900s or Tel Aviv’s modernist open balconies of the 1920s–1930s.

To start with, consider the hierarchical feature group listed in Table 1.
A sufficiently large image corpus, labeled with some or all of these features, requires an image classification model (Figure 1). In this model, features exist in a hierarchy, from the individual window, to the facade. Semantically, each feature may be significant as an aspect of the city studied by the researcher. For example, a study concerned with the influence of certain modifications to building by-laws concerning balcony arrangements requires identifying buildings in the city designed with such arrangements.

At the first stage, senior scholars at PI level identify the balcony arrangement feature that constitutes a manifestation of the studied phenomenon in building facades. At the second stage, trained domain experts at MS and PhD levels tag a sufficiently large, but far from exhaustive, dataset of selected buildings in which the desired features exist, producing the training dataset. At the third stage, a CNN model can be trained using this dataset to say whether or not the feature exists in other facades which are not part of the training dataset. At the fourth stage, CNN-produced tagging is examined and analyzed by the PIs, contributing to rearticulation of the semantic swatches and updating the model. The standard procedures relating to the separation of training, testing, and validation datasets establish how well the model performs.

In this way, the research group produces a collection of trained models for the selected features (or feature set). The features range from facade level (whether or not the facade is symmetrical), to element level (the type of window). Once this model set has been developed, the whole dataset can be studied using these models. The patterns identified using these models are analyzed by experts at PI level, reassessed, and can then be used to update features or develop new ones. These can be used to develop a more mature understanding of both the dataset and the feature set.

This iterative process (see Table 2) may lead to the identification of new feature sets and lead to new conceptualizations of the dataset. New features may also become apparent as composites of existing features. If the facades are geocoded—each facade is associated with a street address—this will enable a neighborhood scale analysis of the dataset as well.

### Dataset preparation

Preparing the dataset for training the model is a significant labor-intensive challenge in our proposed method. Significant challenges include selecting and labeling the images for the training dataset. In image classification tasks, developers of CNN designs use two broad approaches. Initial datasets, which are relatively small in size, are prepared in-house such as the AlexNet model (Krizhevsky et al., 2012). Large-scale datasets, such as the ImageNet dataset (Deng et al., 2009) or the Places dataset (Zhou et al., 2014) use crowdsourced methods such as Mechanical Turk. These approaches are suitable where the dataset is a

| Feature type          | Categories                                      |
|-----------------------|-------------------------------------------------|
| Window type           | Shuttered, paned, arched, sliding, arched, dormers |
| Balcony type          | Open, enclosed, recessed, corner, double height  |
| Cladding type         | Tile, sandstone, whitewash                      |
| Window arrangements    | Strip, individual, paired                       |
| Balcony arrangements   | Staggered, inline, cornered                     |
| Facade type           | Symmetrical, asymmetrical                       |
| Stilts                | Complete, partial, absent                       |
| Penthouse             | Complete, partial, absent                       |

A list of hierarchical semantic features or “semantic swatches” and their categories.
In our proposed methodology, the development of the semantic swatches and the labeling of the dataset require specialized domain expertise. This is undertaken by (1) specialist architectural historians who are not only well versed with the historical and architectural concepts at stake, but are also familiar with the city in question, and (2) lay-experts at MS and MArch levels who are versed in the task of reading and analyzing building facades.

For our purpose, one of the most promising aspects of CNNs is that unlike the more established CBIR-based approaches, they do not require explicit feature engineering. In other words, the semantic mapping is produced purely by applying a label to an image, as opposed to having to specify rules about color or texture patterns or types, as was the case in the general-purpose dataset in which the volume of labeled data points ultimately numbers in the millions, where no domain-expert analysis of the object and turk training as described here was conducted (Kovashka et al., 2016).

In our proposed methodology, the development of the semantic swatches and the labeling of the dataset require specialized domain expertise. This is undertaken by (1) specialist architectural historians who are not only well versed with the historical and architectural concepts at stake, but are also familiar with the city in question, and (2) lay-experts at MS and MArch levels who are versed in the task of reading and analyzing building facades.

For our purpose, one of the most promising aspects of CNNs is that unlike the more established CBIR-based approaches, they do not require explicit feature engineering. In other words, the semantic mapping is produced purely by applying a label to an image, as opposed to having to specify rules about color or texture patterns or types, as was the case in the

**Figure 1.** An example of an image with some highlighted features of interest.

**Table 2.** An iterative method for inquiry into an image corpus.

| Task # | Task                                      | Competency                        | Next task |
|-------|-------------------------------------------|-----------------------------------|-----------|
| 1     | Identify semantic features                | Historians                        | 2         |
| 2     | Create datasets for chosen semantic features or swatches | Historians, Grad. Students.       | 3         |
| 3     | Train models                              | Computing specialists              | 4         |
| 4     | Analyze the full corpus using trained models | Historians and computing specialists | 5         |
| 5     | Update features                           | Historians                        | 2         |
Streetscore study or Doersch’s work on Paris. It may well be the case, for example, that some of the views in the StreetView dataset involve trees, buses, cars, street signage or other elements of the landscape obscuring the view of the feature. This will have an effect on the efficiency of training the model, but it does not represent a conceptual impediment in the approach.

Limitations and pitfalls

The interdisciplinary nature of the proposed project provides significant potential rewards as well as a few foreseeable pitfalls. First, the design and training of CNN models involves an iterative process of designing and testing. This is inherently heuristic, and while current models are known to be accurate about 95% of the time, currently it is unclear that these models will definitely advance to a level where they predict the existence or absence of a feature perfectly. The potential for false positive predictions remains, and the training and testing process is expected to unearth several such examples. Second, the proposed research is inherently inductive and not deductive. It has the potential to expand the capacity of architectural historians—who are the primary domain experts—to interrogate large corpuses of pictorial data. It is not intended to be, and will not work in the form of, an automated history machine. The proposed research offers an incremental advancement to the existing methods of architectural history rather than a fundamental transformation in research methods.

Summary

By enabling an iterative process of inquiry over the dataset, the proposed method uses the capabilities of CNN models to turn large image corpuses into an intelligent representation system. While feature sets can be updated, the data set itself cannot be modified.

The proposed method is extensible in several ways. While this example considers building facades, other elements in a dataset of building facades extracted from the Google StreetView record can also be used—such as foliage, street signage, zoning, and other relevant features. The CNN model imposes no inherent limits to such extensibility beyond the necessity for producing a sufficiently large training dataset. This ability to test feature sets is a novel contribution of the proposed methodology and promises to extend the toolset available to scholars of the built environment, posing historical or contemporary research questions.

Proof of need: Tel Aviv Historical Urban Landscape

Introduction

We describe an ongoing research of Tel Aviv’s urban history requiring and employing the method described in the previous section. Tel Aviv’s modernist urban history of the 1920s–1930s is identified with its modernist legacy as a UNESCO world heritage site. In 2011, UNESCO accepted recommendations for a more inclusive approach to the analysis of HUL. HUL is an urban area “understood as the result of a historic layering of cultural and natural values and attributes, extending beyond the notion of ‘historic center’ or ‘ensemble’ to include the broader urban context and its geographical setting” (UNESCO, 2012, p. 3). Such an analysis demands the capacity to examine large corpuses of images, texts, drawings, and other sources of visual culture.

While preservation in Tel Aviv does identify and list iconic modernist buildings, the city’s modernist built environment is arguably a landscape where listed buildings are embedded among non-listed modernist buildings in the urban fabric (Gottesman and Hoffmann, 2019; Metzger-Szmuck, 2004). Veldpaus, Pereira Rodgers, and Colenbrander have shown that the
HUL approach has historical roots in the work of Tel Aviv’s masterplan planner Sir Patrick Geddes (Pereira Rodrigues and Bandarin, 2019; Veldpaus et al., 2013). An earlier study analyzed the urban development of Tel Aviv through a study of the historical morphology of its urban clusters (Benguigui et al., 2006).

HUL is therefore highly relevant for the city of Tel Aviv. The Tel Aviv Preservation Department set out to assess the urban heritage of the city’s 1980s–1990s “post-modern” period beyond specific iconic buildings which depart from the modernist principles. Seeking a HUL-based approach to identifying the large-scale characteristics of the city’s period architecture, the Tel Aviv Preservation Department commissioned a research grant to this study.

The case study we report in this section consists of a “proof of need” for urban research of large datasets for questions of urban heritage, relevant for many cities worldwide. The method proposed in this paper, and described in the previous section, plays an important role in developing an HUL-based approach to the study of Tel Aviv’s post-modern period. This study is not only of relevance to this city, it will also allow collaboration with additional cities. Adapting the digital catalog of each city and enabling identification and theorization of each phenomenon, it will produce an ever-growing catalog useful for understanding diverse urban phenomena. The details of this work are discussed below.

**Developing the swatches, dataset, and training models**

This work is an interdisciplinary collaboration between architectural and urban historians and computing specialists who train, test, and validate the neural network models. The architectural historians involved are specialists in architectural history of the 20th century, specifically of the city of Tel Aviv (Allweil, 2016a, 2016b; Allweil and Zemer, 2019). These experts designed the swatches. PI and post-doc experts in Tel Aviv’s architectural and urban history identified a series of related swatches. Afterwards, MS and MArch level experts in architectural and urban history conducted dataset preparation tasks, which consist of labeling and verifying the training data under the supervision of the principal experts. The trained dataset consists of roughly 5000 expert-tagged images.

Instead of using randomly selected building facades for inclusion in the training dataset, we developed a web-based interface in which students can look up building facades by street address, tag it for the appropriate semantic swatches and write it to a database (see Figure 2). We use the Google StreetView API to access the image corpus of Tel Aviv street facades.

The approach is to select the identified urban blocks to demonstrate one of various semantic features modeled in each swatch, which are later applied to other urban blocks.

We developed the preliminary set of semantic swatches to reflect changes in the building by-laws which shaped the design of buildings of the 1980s in Tel Aviv through changing aesthetics and density requirements. To take one example, for a period of time, buildings could not have stacked balconies; they had to be staggered. Staggered balconies were consequently identified as a semantic swatch. Students identified buildings in Tel Aviv which show this feature, labeled these, and added these to the training dataset. Using a standard data augmentation technique from the practice of machine learning modeling, multiple images of each building were added using different field-of-view, pitch and heading, artificially increasing the size of the training set by generating many realistic variants of each training instance (Géron, 2017: 465). This serves as a regularization technique and prevents the model from overfitting to the data, resulting in a 12,500 image dataset.

This data collection process is labor intensive and is ongoing. Wherever a sufficiently large dataset has been completed, a model is trained using the standard practices of machine learning. We use the Keras machine learning library for this purpose.
Training instance (G) socially increasing the size of the training set by generating many realistic variants of each data augmentation technique from the practice of machine learning modeling, multiple subsequently identified as a semantic swatch. Students identified buildings in Tel Aviv which could not have stacked balconies; they had to be staggered. Staggered balconies were a reaction to aesthetic and density requirements. To take one example, for a period of time, buildings by-laws which shaped the design of buildings of the 1980s in Tel Aviv through changing architectural and urban history conducted dataset preparation tasks, which consist of label-history identified a series of related swatches. Afterwards, MS and MArch level experts in architectural historians involved are specialists in architectural history of the 20th century, and computing specialists who train, test, and validate the neural network models. The This work is an interdisciplinary collaboration between architectural and urban historians. Developing the swatches, dataset, and training models understanding diverse urban phenomena. The details of this work are discussed below.

This study is not only of relevance to this city, it will also allow collaboration with additional cities. Adapting the digital catalog of each city and enabling identification and the-torialization of each phenomenon, it will produce an ever-growing catalog useful for

Figure 2. Dataset preparation using Google Google StreetView.

Using the models and updating swatches

The trained models are applied to a geotagged set of building facades for Tel Aviv. This gives an indication as to the neighborhoods in the city where each semantic feature is evident. In other words, it indicates the nature of the presence of the feature in the urban landscape. Various processes make a HUL-based approach more feasible to historians: the iterative process of identifying swatches, creating a dataset for each of these swatches, using trained models to study the urban landscape through a geotagged corpus of facade images, and updating existing swatches or developing new ones. The possibilities, limits, and implications of this new method are discussed in the next section.

The last stage of the ongoing project involves contributing the findings of our approach to the existing GIS database maintained by the city of Tel Aviv.

Discussion

Implications for the archive

A method enabling historians to take examples and use them to systematically identify a larger set of similar examples has significant implications for the concepts of the archive and the canon itself. The archive does not merely maintain a record of its subject, it “produces” it (Derrida, 1996). Further, the archive shapes what questions historians ask (Farge and Davis, 2013).

Archives of the built environment have typically consisted of collections of drawings, models, photographs, slides, sketches, and other pictorial or three-dimensional objects in addition to contracts, letters, and other documents. Such documentation tends to be available for prominent or important buildings and has shaped the architectural canon.

Date and Allweil

Figure 2. Dataset preparation using Google Google StreetView.
Nevertheless, it does not enable posing systematic research questions regarding architectural and urban processes (Fletcher, 1931; Kostof, 1985). As seen from the standard survey course offered in architectural programs around the world, the mainstream approach to introducing the history of architecture relies on this architectural canon. A HUL-based approach to history and heritage enabled by the method proposed in this paper has the potential to augment our existing capacity to develop case studies. Consequently, it has the potential to enrich the study of architectural history.

We expect this capability to be of great use to architectural and urban historians and will help them study the city in ways which have hitherto been unavailable to them. Such an analysis is expected to contribute towards a revision of the chosen swatches by indicating new features which could be of interest. New datasets will be developed for these semantic features, and the training processing described above will be repeated.

**Limits, future work, and conclusion**

This paper proposes a method using recent advances in computer vision to enable architectural and urban historians to take advantage of large-scale image corpuses of built environments that are now ubiquitously available. Our method expands the ways in which scholars can study architectural and urban history, making a HUL-based approach to history and urban heritage more feasible.

The method proposed in this paper is currently being tested in the case study described in Section “Proof of need: Tel Aviv Historical Urban Landscape,” whose specific findings are the subject of another paper. The current state of the project indicates that there is potential to extend the dataset production process in an encyclopedic direction toward developing a stable hierarchy of composable systems of semantic swatches. This methodological approach involves codifying the expert knowhow (which currently underlies the dataset production process) so that dataset production can be scaled up via crowdsourcing. We anticipate that such a shift to a crowdsourcing approach will involve further conceptual innovations which lie beyond the present scope of work.

The iterative process in the proposed model is heuristic. The interdisciplinary nature of the participants means that this involves the historians developing high level intuitions about the CNN models. As indicated in Section “Limitations and pitfalls,” the role of the model is to assist historical scholarship by enabling a new kind of inquiry into a large-scale image corpus—expanding the tools of scholars rather than replacing them with a “historical machine.”

Our methodology introduces new questions for urban scholars, expanding the scope of the field of urban and architectural history and related fields of urban studies. We propose the applicability of our method to study other questions for which other types of image corpuses such as streetscapes or residential floor plans will be appropriate.

**Acknowledgements**

The authors would like to acknowledge the support from the Research & Development Center at the Faculty of Architecture & Town Planning, Technion, Israel Institute of Technology, the Preservation Department of the Tel Aviv Municipality.

**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: The authors would like to acknowledge the support from the Israel Science Foundation (ISF) for its support through grant #2029064.

ORCID iD
Kartikeya Date https://orcid.org/0000-0002-0226-9527

Note
1. For the purpose of this paper, the terms picture and its derivatives—pictorial and image—refer to a two-dimensional pixel representation in which each pixel has a color defined as some combination of three colors (red, green, and blue).

References
Alexander C (1977) A Pattern Language: Towns, Buildings, Construction. Oxford: Oxford University Press.
Allweil Y (2016a) Anarchist city? Geddes’s 1925 Anarchist housing-based plan for Tel Aviv and the 2011 housing protests. In: White RJ, Springer S and Lopes de Souza M (eds) The Practice of Freedom: Anarchism, Geography and the Spirit of Revolt. Lanham: Rowman & Littlefield Publishers, pp.43–64.
Allweil Y (2016b) Homeland: Zionism as Housing Regime, 1860–2011. Abingdon: Routledge.
Allweil Y and Zemer N (2019) Housing-based urban planning? Sir Patrick Geddes’ modern master-plan for Tel Aviv, 1925. Urban Planning 4(3): 167–185.
Araldi A and Fusco G (2017) Decomposing and recomposing urban fabric: The city from the pedestrian point of view. In: Gervasi O, Murgante B, Misra S, et al. (eds) Computational Science and Its Applications – ICCSA 2017. New York: Springer International Publishing, pp.365–376.
Behnisch M, Hecht R, Herold H, et al. (2019) Urban big data analytics and morphology. Environment and Planning B: Urban Analytics and City Science 46(7): 1203–1205.
Benguigui L, Blumenfeld-Lieberthal E and Czamanski D (2006) The dynamics of the Tel Aviv morphology. Environment and Planning B: Planning and Design 33(2): 269–284.
Berghauser Pont M, Stavroulaki G, Bobkova E, et al. (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(7): 1226–1242.
Bernheimer R (1961) The Nature of Representation: A Phenomenological Inquiry. New York: New York University Press.
Bodenhamer B, David J, Corrigan J, et al. (2015) Deep Maps and Spatial Narratives. Bloomington: Indiana University Press.
Bodenhamer DJ (2013) Beyond GIS: Geospatial technologies and the future of history. In: von Lünen A and Travis C (eds) History and GIS: Epistemologies, Considerations and Reflections. Netherlands: Springer, pp.1–13.
Boy JD and Uitermark J (2017) Reassembling the city through Instagram. Transactions of the Institute of British Geographers 42(4): 612–624.
Bucciarelli LI (1988) An ethnographic perspective on engineering design. Design Studies 9(3): 159–168.
Carpo M (2001) Architecture in the Age of Printing: Orality, Writing, Typography, and Printed Images in the History of Architectural Theory. Cambridge: The MIT Press.
Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, July 21-26, 2017. pp. 1251–1258. Available at: https://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html.
Colaninno N, Roca J and Pfeffer K (2011) An automatic classification of urban texture: Form and compactness of morphological homogeneous structures in Barcelona. Available at: https://www.econstor.eu/handle/10419/120085.

D’Acci L (2019) On urban morphology and mathematics. In: Batty M and D’Acci L (eds) The Mathematics of Urban Morphology. New York: Springer International Publishing, pp.1–18.

Dastin L and Galison P (1992) The image of objectivity. Representations 40: 81–128.

Datta R, Joshi D, Li J, et al. (2008) Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40(2): 1–60.

Deng J, Dong W, Socher R, et al. (2009) ImageNet: A large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition, Miami, FL, 20–25 June 2009, pp.248–255. Piscataway: IEEE.

Derrida J (1996) Archive Fever: A Freudian Impression. Chicago: University of Chicago Press.

Doersch C, Singh S, Gupta A, et al. (2015) What makes Paris look like Paris? Communications of the ACM 58(12): 103–110.

Farge A and Davis NZ (2013) The Allure of the Archives (T. Scott-Railton, Trans.). 1st ed. New Haven: Yale University Press.

Ferguson ES (1994) Engineering and the Mind’s Eye. Cambridge: The MIT Press.

Fletcher B (1931) A History of Architecture on the Comparative Method. London: Batsford.

Frampton K (1981) Modern Architecture: A Critical History. Oxford: Oxford University Press.

Gane N and Beer D (2008) Archive. In: New Media: The Key Concepts. Oxford: Berg Publishers, pp.71–86.

Géron A (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 1st ed. Sebastopol: O’Reilly Media.

Goel V (1995) Sketches of Thought. Cambridge: The MIT Press.

Goldschmidt G (1991) The dialectics of sketching. Creativity Research Journal 4(2): 123–143.

Golledge RG and Spector AN (1978) Comprehending the urban environment: Theory and practice. Geographical Analysis 10(4): 403–426.

Goodfellow I, Bengio Y and Courville A (2016) Deep Learning. Illustrated ed. Cambridge: The MIT Press.

Goodman N (1976) Languages of Art: An Approach to a Theory of Symbols. 2nd ed. Indianapolis: Hackett Publishing Company.

Gottesman R and Hoffmann J (2019) Actual and intangible in Tel Aviv: A reexamination of conservation strategies in a modern city. In: Pereira Roders A and Bandarin F (eds) Reshaping Urban Conservation: The Historic Urban Landscape Approach in Action. New York: Springer, pp.473–482.

Habraken NJ and Teicher J (2000) The Structure of the Ordinary: Form and Control in the Built Environment. Cambridge: MIT Press.

Haken H and Portugali J (2003) The face of the city is its information. Journal of Environmental Psychology 23(4): 385–408.

Harley JB (2001) The New Nature of Maps: Essays in the History of Cartography. Baltimore: Johns Hopkins University Press.

Harris K (2014) [Digital] Archive. In: Ryan M-L, Emerson L and Robertson BJ (eds) Johns Hopkins Guide to Digital Media & Textuality. Baltimore: Johns Hopkins University Press, pp.16–18.

He K, Zhang X, Ren S, et al. (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016, Las Vegas, NV, USA, June 27–30 2016, pp.770–778. Available at: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

Heyman S (2015) Photos, photos everywhere. The New York Times, 29 July. Available at: https://www.nytimes.com/2015/07/23/arts/international/photos-photos-everywhere.html.

Hochman N and Manovich L (2013) Zooming into an Instagram City: Reading the local through social media. First Monday 18(7). https://journals.uic.edu/ojs/index.php/fm/article/view/4711/3698

Ibrahim MR, Haworth J and Cheng T (2019) URBAN-i: From urban scenes to mapping slums, transport modes, and pedestrians in cities using deep learning and computer vision. Environment
and Planning B: Urban Analytics and City Science. Epub ahead of print 6 May 2019. DOI: 10.1177/2399808319846517.

Isola P, Xiao J, Parikh D, et al. (2014) What makes a photograph memorable? IEEE Transactions on Pattern Analysis and Machine Intelligence 36(7): 1469–1482.

Jencks C (2002) The New Paradigm in Architecture: The Language of Postmodernism. New Haven: Yale University Press.

Kalay YE (2004) Architecture’s New Media. Cambridge: MIT Press.

Kang J, Körner M, Wang Y, et al. (2018) Building instance classification using street view images. ISPRS Journal of Photogrammetry and Remote Sensing 145: 44–59.

Kostof S (1985) A History of Architecture: Settings and Rituals (Elyachar Central Library - Store G-4974). Oxford: Oxford University Press.

Kostof S (2000) The Architect: Chapters in the History of the Profession. 1st ed. California: University of California Press.

Kovashka A, Russakovsky O, Fei-Fei L, et al. (2016) Crowdsourcing in computer vision. Foundations and Trends® in Computer Graphics and Vision 10(2): 103–175.

Krizhevsky A, Sutskever I and Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, et al. (eds) Advances in Neural Information Processing Systems 25. New York: Curran Associates, Inc., pp.1097–1105.

Latour B (1986) Visualization and cognition: Thinking with eyes and hands. In: Kuklick H (ed.) Knowledge and Society: Studies in the Sociology of Culture Past and Present. Vol. 6. Stanford: Jai Press, pp.1–40.

Lavee M (2017) Digital humanities and crowdsourcing in the Cairo Geniza. Deot 81: 40–44.

Lavee M (2019) Tikkoun Sofrim – Combining HTR and Crowdsourcing for Automated Transcription of Hebrew Medieval Manuscripts. DH2019, Utrecht University, Utrecht. https://dev.clariah.nl/files/dh2019/boa/0568.html

LeCun Y, Boser B, Denker JS, et al. (1989) Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4): 541–551.

LeCun Y, Bottou L and Haffner P (1998) Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11): 2278–2324.

Leszczynski A (2019) Digital methods II: Digital-visual methods. Progress in Human Geography 43(6): 1143–1152.

Long J, Shellhamer E and Darrell T (2015) Fully convolutional networks for semantic segmentation. ArXiv:1411.4038 [Cs].

Lynch K (1960) The Image of the City. Cambridge: MIT Press.

Lynch M and Woolgar S (eds) (1990) Representation in Scientific Practice. Cambridge: MIT Press.

Manoff M (2004) Theories of the archive from across the disciplines. Portal: Libraries and the Academy 4(1): 9–25.

Metzger-Szmuck N (2004) Dw ellings on the Dunes. Editions de l’eclat. Paris.

Milligan I (2016) Lost in the infinite archive: The promise and pitfalls of web archives. International Journal of Humanities and Arts Computing 10(1): 78–94.

Morris AEJ (1979) History of Urban Form before the Industrial Revolutions (Architecture & Town Plann. - General Collection 711.4 MO; 2nd ed.). New York: Wiley.

Naik N, Kominers SD, Raskar R, et al. (2017) Computer vision uncovers predictors of physical urban change. Proceedings of the National Academy of Sciences 114(29): 7571–7576.

Naik N, Philipoom J, Raskar R, et al. (2014) Streetscore—Predicting the Perceived Safety of One Million Streetscapes, pp.779–785.  https://www.cv-foundation.org/openaccess/content_cvpr_workshops_2014/W20/html/Naik_Streetscore_-_Predicting_2014_CVPR_paper.html

Oliveira V (2016) Urban Morphology. New York: Springer International Publishing.

Pereira Rodgers A and Bandarin F (eds) (2019) Reshaping urban conservation. In: Reshaping Urban Conservation: The Historic Urban Landscape Approach in Action. Berlin: Springer, pp.3–20.

Perez J, Fusco G, Araldi A et al. (2018, September). Building typologies for urban fabric classification: Osaka and Marseille case studies. In: International conference on spatial analysis and modeling (SAM), Tokyo, Japan, September 2018. https://hal.archives-ouvertes.fr/hal-02176599
Kartikeya Date, PhD, is an architect from Mumbai India and is currently a post-doctoral fellow at the Technion, Israel Institute of Technology, Haifa, Israel. He completed his PhD in design theories and methods at UC Berkeley.

Yael Allweil, PhD, is an architect and Associate Professor in the Faculty of Architecture and Town Planning at the Technion, Israel Institute of Technology, Haifa, Israel, where she heads the HousingLab: The History and Future of Living. She completed her PhD in architecture history at UC Berkeley exploring the history of Israel–Palestine as a history of the gain and loss of citizen housing. Her research was published in the monograph Homeland: Zionism as Housing Regime 1860–2011 (Routledge, 2017) and several journal articles in Urban Studies, Footprint, Architecture Beyond Europe, City, TDSR and IJIA.