“Reading” cities with computer vision: a new multi-spatial scale urban fabric dataset and a novel convolutional neural network solution for urban fabric classification tasks

Zhou Fang
University of Cambridge
United Kingdom
zf242@cam.ac.uk

Jiaxin Qi
Nanyang Technological University
Singapore
jiaxin003@e.ntu.edu.sg

Tianren Yang
University of Cambridge
United Kingdom
ty290@cam.ac.uk

Li Wan
University of Cambridge
United Kingdom
lw423@cam.ac.uk

Ying Jin
University of Cambridge
United Kingdom
yj242@cam.ac.uk

ABSTRACT

This paper builds on the proven track record of CNN-based pattern recognition and feature extraction methods, and reports a novel model that classifies urban fabric samples of metropolitan areas in terms of (1) which city they belong to, (2) what types of urban fabric they belong to, and (3) which historic period they originate from. Currently, such tasks require intensive manual work by senior professionals, and even then, inconsistencies and errors occur. Our work is based on a novel urban fabric dataset of four metropolitan areas with distinct typologies (linear development, open block, gated compound, medieval region, irregular grid and orthogonal grid), which consist of high resolution 3-dimensional built form data and hierarchical street networks. The classification model presented in this paper is the first that is capable of predicting the city origin, urban fabric pattern type and construction period. The novelty is also characterised by jointly considering urban fabric features across multiple spatial scales. The experiments demonstrate that this multi-scale approach can capture a wide range of urban fabric features across cities, urban fabric pattern types and development periods. We further find that the effectiveness can be enhanced by appending an auxiliary network for identifying the most appropriate combinations of the multiple spatial scales in line with the classification task. The dataset and model can massively scale up the productivity of researchers and professionals working on cities.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision tasks; Supervised learning**; • **Applied computing** → **Architecture (buildings)**.

KEYWORDS

Urban fabric classification, Computer vision, Multi-spatial scale, Supervised learning, Multi-task learning

ACM Reference Format:

Zhou Fang, Jiaxin Qi, Tianren Yang, Li Wan, and Ying Jin. 2020. “Reading” cities with computer vision: a new multi-spatial scale urban fabric dataset and a novel convolutional neural network solution for urban fabric classification tasks. In *28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '20)*, November 3–6, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3397536.3422240>

1 INTRODUCTION

Understanding the typology of cities and their evolution paths is a fundamental topic in urban planning, design and geography. Generations of scholars and practitioners aspire to grapple with the increasing variety and complexity of urban morphology. Since the 18th century, a vast modern literature about urban form and urban process have been published and widely spread. From the *1728 Map of Rome*, by Giambattista Nolli, *The Art of Building Cities*, by Camillo Sitte, down to the Kevin Lynch’s work *The Image of The City*, generations of scholars have used their professional eyes to “read” the colourful and diversified physical urban forms, also known as urban fabric, and tried to answer a classic research question in urban studies: How to “read” urban fabric and how to scrutinize urban configurations through it [11]. Since the application of typology concept in architectural studies was firstly proposed, urban fabric typology studies, also known as urban typo-morphology studies, has engaged the attention of researchers over time [13].

Along with great effort made by urban typology scholars, architects, planners and researchers have benefited profoundly from the urban fabric typology studies. By studying existing urban fabric, professionals are inspired to incorporate the classic urban fabric patterns into their own design. At the same time, urban fabric typology studies also facilitate other fields of studies. As urban fabric reflects cities’ micro and macro spatial structure and influences human behaviours, it sheds a light on understanding the interaction between the built environment (e.g. street blocks, transport network) and the social-economic development of cities. The strong correlations discovered between urban fabric and land use changes explained how a city grows and experts have used this to predict

future urban expansion patterns [9]. [18] presents their findings about heat-energy demand of different urban fabric types and provides insightful suggestions for achieving environmentally-friendly development through innovative urban design strategies. Given its significance and wide range of usage, urban fabric typology studies are in need of keep stepping forward.

Start from 1748, when the Nolli Map was presented, scholars have begun to read and interpret urban fabric using figure-ground images and abstract scratches [15]. This traditional, conceptual and intuitive description of urban fabric well preserves the spatial configuration of urban fabric and helps professionals to sense and distinguish different fabric types through intuitive way. However, the process of producing an image with professional urban fabric typology mapping requires massive manual labelling works, and the quality of the mapping largely depends on the observer's expert knowledge and personal cognition. These make the traditional approach has several limitations in practice. First, the workload of labelling would become prohibitive when the spatial scale of the study expands to large metropolitan areas. Second, the underlying subjectivity in typology labelling may result in difficulties in replicating the studies across cities and over time. Recently, researchers have started to use parametric approaches to interpret urban fabric. Series of numerical indicators concisely describing urban fabric are selected and clustered with automatic classification algorithms. This quantitative method avoids observers' personal cognition and the parametric representation of urban fabric makes the automatic urban fabric clustering become feasible and easy to reproduce. Nevertheless, because the geometric information and spatial configuration of the fabric are hard to be fully represented using a few numerical indicators, the experience based choice of limited indicators may lead the automatic urban fabric clustering process into biased directions [4].

To overcome the main limitations of existing urban fabric classification approaches and provide a fully automated, spatial configuration sensitive classification solution, we propose to introduce state of the art computer vision technology into the field of urban fabric typology studies, targeting to equip computers with urban planners' ability to interpret, understand and classify diverse urban fabric. Inspired by the classic computer vision tasks: image classification and object detection, we integrate 3-dimensional built form data and hierarchical road network data to form image-like data patches and use this pictorial representation of urban fabric to train a Convolutional Neural Network (CNN) based model towards ground truth fabric type labels provided by professional urban planners. Although CNN based approach has shown decent spatial pattern recognition capabilities in other computer vision tasks, we intend to make two modifications in both the dataset and the CNN model to address the specific needs of our urban fabric classification tasks. First, considering the spatial scale of urban fabric features can vary significantly across areas of different development periods in one city and also across cities, and the potential influence of different extents of surrounding context upon urban fabric classification result [22], we use a set of image-like data patches of varying spatial scales to pictorially represent each urban fabric. This multi-spatial scale dataset provides an opportunity for our model to perform classification by jointly considering urban fabric at multiple spatial scales. Second, given the evidence of the close correlations between

distinct urban fabric patterns and urban formation periods [11], together with the aim of extending the features of our urban fabric classification model, we extend our model to one that is capable of simultaneously handling three different classification tasks: predicting urban fabric's city origin, pattern type and formation period.

More specifically, we first present a novel multi-spatial scale urban fabric dataset, which includes about 66k multi-scale urban fabric samples extracted from four cities in Europe: Amsterdam, Netherlands, Berlin, Germany, Barcelona, Spain and Prague, the Czech Republic. 2-dimensional data are enhanced by high resolution Light Detection and Ranging (LiDAR) data and OpenStreetMap (OSM) data, which capture detailed 3-dimensional built form information and hierarchical road network information. We also present our systematic process for creating the dataset and list the details of used source data, so that the community can reproduce it or create a new one for their own use. Then we propose a CNN based urban fabric classification model to simultaneously predict urban fabric's origin, pattern type and formation period by jointly considering urban fabric in multiple spatial scales. Our results show that our urban fabric classification model can achieve over 0.90, 0.70 and 0.70 F1-Scores on the three classification tasks respectively. The necessity of jointly considering urban fabric in multiple spatial scales on urban fabric classification tasks is verified through an ablation test. Further, we test the generalization capability of our model through a designed experiment. Although the performance of the model drops when handling absolute new urban fabric patterns, we propose a transfer learning solution that can significantly boost model prediction capability by only providing a small amount of additional labelled training data.

The main contributions of this work are therefore as follows:

- This work introduces CNN, one of the state-of-the-art computer vision technique, into the field of urban typomorphology and the convincing results from this interdisciplinary study suggest the great potentials of the further deployment of CNN based approaches in the field of urban studies.
- This work provides a new way to pictorially represent the 3-dimensional urban fabric. By integrating high resolution 3-dimensional built form and hierarchical road network data and forming a multi-channel image-like data patches, the spatial configuration of urban fabric can be well maintained while the data samples are compatible with CNN model.
- A large labelled multi-spatial scale urban fabric dataset, containing 66k urban fabric samples from four cities in Europe, is set up and the process of creating the dataset is also presented in detail, so that the community can reproduce and create a new one for their own use.
- A CNN based urban fabric classification model is proposed. To the best of our knowledge, this is the very first that is capable of predicting the city origin, urban fabric pattern type and construction period. Its novelty is also characterised by jointly considering urban fabric features across multiple spatial scales.
- A transfer learning based approach aiming to boost the model's generalization capability is presented and this allows

the model to be directly utilized for urban fabric classification tasks in new cities with modest requirement on extra training data.

For the rest of this paper, we first provide an overview of related works in Section 2. The process of creating our proposed multi-spatial scale urban fabric dataset, and its statistic characteristics are discussed in Section 3. The urban fabric classification model is presented in Section 4 followed by experiments and results in Section 5. Section 6 introduces a transfer learning based approach that allows the direct use of the proposed model for urban fabric type and formation process mapping. Finally, Section 7 concludes the paper.

2 RELATED WORKS

In most previous urban fabric typology studies, the classification of urban fabric types was performed manually by urban scholars using their disciplinary knowledge and experience. Figure-ground image and abstract scratch were widely used to pictorially represent the geometry of the urban fabric and the classification results were always shown together with the illustration diagrams. In Spiro Kostof’s book, *The City Shaped*, the author discussed patterns and elements of urban fabric in a historical perspective. Through “reading” historic drawings and maps, the renowned and representative urban forms in history are systematically recognised as “organic pattern”, “grid”, “diagram” and “grand manner” regarding not only their spatial configuration but also historical background [11]. Although this classic approach offers an intuitive way to cluster urban fabric, its limitations are obvious: (1) mounting workload for manual labelling and (2) underlying subjectivity in typology labelling.

Recently, researchers have begun to interpret urban fabric in quantitative ways. Series of numerical indicators are selected to parametrically represent urban fabric. In [2], 64 standardized urban morphology indicators are summarized for three spatial scales of analysis, including “area”, “floor area”, “building perimeter” for building scale, “block net compacity”, “area of holes in a block”, “ratio between the holes’ area and the block’s area” for block scale, and “number of buildings in the RSU” and “distance to the city center” for the Reference Spatial Unit (RSU) scale. These indicators are either collected through land survey or derived based on rules defined by researchers. Apart from these, other specific indicators are also considered in some urban fabric linked studies. “Imperious fraction”, “Sky Exposure Factor (SEF)”, “Sky View Factor (SVF)”, these are some indicators always considered in micro-climate related studies [3, 5, 17]. Other indicators, such as “walkability index”, “bikeability index”, “distance to station”, etc. are normally included in transportation related studies [7]. In general, with parametric representation of urban fabric, the classification process can be automated by using machine learning algorithms, for example Regression Trees (RTs), Random Forest (RF) and Support Vector Machines (SVMs), etc. [2] showed the feasibility of broadening the geographic extent of research into national-wide using standardized urban indicators and RTs algorithm. However, the detailed geometric information is hard to be represented numerically, which means limited numbers of indicators can not capture a wide range of complex spatial configurations of urban fabric and the clustering result cannot be sensitive to that as well.

Given the proven pattern recognition and feature extraction capability, CNN has been applied in many fields of studies. However, urban studies, especially urban typo-morphology field, rarely apply this technique and the limited attempts of using CNN in urban morphology studies were far from fully utilizing the capability of CNN and have not introduced a proper way to incorporate domain knowledge. [16] used a Convolutional Autoencoder (CAE) to extract spatial features of street network and used self-organizing map to self-cluster similar patterns regarding the encoded dense vectors. Similar methodology has been adopted for urban fabric classification in Nan Jing, China [4]. These studies confirmed that trained CNN model is able to encode the pictorial representations of the urban forms to vector based ones, while it considers the overall spatial structure and configuration. However, the self-clustering process adopted, which is an unsupervised learning process, did not provide any opportunity to involve domain knowledge to guide the classification process. This turns the clustering process into a “black box” and the classification results need careful interpretations before further use. Moreover, only figure-ground images are used to represent urban fabric in [4]. The complex 3-dimensional spatial features of urban fabric are not recorded in the data samples and extracted by CAE. This makes the classification process only consider 2-dimensional features rather than a more complex 3-dimensional one. Last but not least, the image samples used to train CAE in previous studies are collected in a fixed spatial scale. Considering (1) the significance variation of spatial scale of generic urban fabric features across areas of different development periods in one city and also across cities, and (2) the potential influence of different extents of surrounding context upon urban fabric classification result [9], there is a great need to upgrade existing model to a multi-spatial scale adaptive one, which can jointly consider urban fabric features across multiple spatial scales.

Our proposed solution targets to address the limitations and challenges identified in literature. First, we turn the self-clustering process proposed in [16] and [4] into a supervised one, where the model can learn how professionals classify the urban fabric from typology labels provided by professional urban planners. Besides, we set up a novel multi-spatial scale, 3-dimensional urban fabric dataset, and use it to train and test the proposed classification model which is capable to classify urban fabric samples of metropolitan areas in terms of (1) which city they belong to, (2) what types of urban fabric they belong to, and (3) which historic period they originate from. Lastly, we extend the proposed model by arming it with the ability to intelligently select the most appropriate combination of spatial scales in line with classification task.

3 DATASET

In this section, we describe the overall process to set up the proposed multi-spatial scale urban fabric dataset, which contains 66k multi-spatial scale urban fabric samples extracted from four European cities: Amsterdam, Netherlands, Berlin, Germany, Barcelona, Spain and Prague, the Czech Republic, with generic urban fabric patterns and broad development periods labelled out for training and validation purposes. An overview of the process, which consists of two stages: multi-channel base map preparation and multi-spatial scale data sample collection, is shown in Figure 1 and detailed description of each stage is provided in the following sub-sections.

Figure 1: Overall process to set up multi-spatial scale urban fabric dataset

3.1 Multi-channel base map preparation

The base map used to set up the new urban fabric dataset is prepared with a series of urban fabric related elements, including 3-dimensional built form, hierarchical street network and street boundary defined land parcels. We use three overlapping layers, or “channels,” of an image to pictorially represent the comprehensive 3-dimensional urban fabric.

High resolution 3-dimensional built form data (channel 1): To set up this channel, we first collect multi source data, including open-source vector representation of building footprints provided by OSM, public accessible Digital Elevation Model (DEM) data from the Shuttle Radar Topography Mission (SRTM) and high resolution Digital Surface Model (DSM) data delivered by LiDAR system (Amsterdam: Altitude File Netherlands 3 (AHN3) point clouds data distributed by Public Service On the Map (PDOK) platform, Berlin: 3D building models data in Level of Detail 1 (LoD 1), provided by the Central Office for House Coordinates and House Rings (ZSHH), Barcelona: LiDAR 2nd Coverage (2015-present) point clouds data, provided by National Geographic Institute (IGN), Prague: Digital Surface Model of the Czech Republic of the 1st generation (DMP 1G), provided by CUZK Land Survey Office). Then, project these data to the local geo-coordinate system, raster, scale and process them into three raster images, Building Footprint Image (BFI), Digital Elevation Image (DEI) and Digital Surface Image (DSI), sharing the same extent, geo coordinate system and resolution. Finally, the three-dimensional built form data (3D_BFI) can be derived as: $3D_BFI = BFI \odot (DSI - DEI)$, where $\odot$ is the pixel-wise multiplication, and the diagrammatic illustration of this process is shown in Figure 2.

Hierarchical street network data (channel 2): Open-source vector representation of street network with street level attributes provided by OSM is used to create the second channel in base maps.

Figure 2: Process to derive 3-dimensional built form data (3D_BFI)

Links (represent streets) acquired from OSM in “motorway”, “trunk”, “primary”, “secondary”, “tertiary”, and “residential” categories are projected and rasterised with different line thicknesses, representing different street hierarchies. In this channel, pixels stored value 1 represent the presence of streets while 0 value pixels represent the absence of streets.

Street boundary defined land parcels (channel 3): In the third channel, we define the boundary of each land parcel within the base map and assign identification number to it, which will be used to generate attention mask that highlights the land parcel located at the centre (the one to perform classification). To produce this channel, we polygonise the hierarchical street network data

Figure 3: Examples of six generic urban fabric typologies considered in this work and their spatial distributions across the four European cities. Here points in different colors represent urban fabric samples categorised in different classes: blue-“linear development”, green-“openblock”, red-“gated compound”, yellow-“medieval region”, brown-“irregular grid” and cyan-“orthogonal grid”.

(channel 2) and assign unique numbers to identified polygons (represent land parcels) followed by rasterising the polygons back to raster image with identification numbers being used for burn-in values.

With the three channels prepared, we concatenate them to form a base map and pass it to a group of urban planning experts to label the ground truth urban fabric classification results for each street boundary defined land parcel based on their experience and domain knowledge.

3.2 Multi-spatial scale data sample collection

In the data sample collection stage, shown in Figure 1, we first crop three multi-channel data patches of size $128\text{px} \times 128\text{px}$, $256\text{px} \times 256\text{px}$ and $512\text{px} \times 512\text{px}$ at each randomly selected location from the prepared multi-channel base map, representing three spatial scales which are frequently adopted in urban studies: lots scale(128m), island scale(256m) and fabric scale(512m) [1]. Smaller scales, building scale(64m), is not considered because most of the land parcels in our dataset are larger than 64m by 64m area. As to larger spatial scales(1024m), we excluded it because with a few number of 1024m by 1024m data patches fed in, the model will be easily overfitted for classification task 1 by remembering the whole city map rather than focus on the urban fabric and the city identities behind. Then, we replace the third channels (street boundary defined land parcels) with binary attention masks, which highlight the land parcel located at the centre. After the three patches in different sizes are all resized to $128\text{px} \times 128\text{px}$, we stack them to form a multi-spatial scale data sample. We then assign the labels for the highlighted center land parcel as the ground truth classification results for the sample. The non-highlighted parts of the sample are seen as the context provided for the classification of the highlighted

center land parcels. With above sampling and labeling strategy, the samples with identical highlighted center land parcels and sharing overlapping regions are seen as data augmentation.

3.3 Urban fabric dataset for Amsterdam, Barcelona, Berlin and Prague

Using the process described above, we created a multi-spatial scale urban fabric dataset for four European cities: Amsterdam, Netherlands, Berlin, Germany, Barcelona, Spain and Prague, the Czech Republic. These four cities were selected not only for their rich mixture of different fabric pattern types but also because of their representative city expansion processes in Europe. The cities began from a medieval core, grew in carefully designed grid regions, and expanded to surroundings with irregular grid patterns and linear development patterns at the urban fringe.

The dataset contains about 66k multi-spatial scale urban fabric data samples and each sample come with three ground truth labels for three classification tasks: (1) which city the sample belongs to, (2) what types of urban fabric it belongs to, and (3) which historic period it originates from. All samples are randomly extracted from a 10km by 6km rectangular region in the four cities by adopting same sampling density. Across the dataset, total six distinct urban fabric typologies (linear development, open block, gated compound, medieval region, irregular grid and orthogonal grid) and four development periods (before 1850, 1850 to 1900, 1900 to 1950, post 1950) are identified by urban planning experts based on their disciplinary knowledge, land survey and historic maps. Examples of six generic urban fabric typologies extracted from four cities are summarised in Figure 3 together with their spatial distributions in the four cities.

Figure 4 provides more details on characteristics of the dataset. To our knowledge, the dataset we set up in this work is the largest

labelled 3-dimensional urban fabric dataset available for classification tasks to the community, in terms of the number of samples, number of samples per urban fabric type category, the number of urban fabric categories, as well as the number of cities covered in the dataset. The dataset will be publicly available after finishing the work.

Figure 4: Characteristics of the multi-spatial scale urban fabric data set for Amsterdam, Barcelona, Berlin and Prague

4 URBAN FABRIC CLASSIFICATION MODEL

In this work, we train models that take multi-spatial scale urban fabric data samples as input and output three sets of the probability on each classification categories for the three classification tasks: (1) which city they belong to, (2) what types of urban fabric they belong to, and (3) which historic period they originate from. In this section, we first present a straight-forward solution to handling the multi-spatial scale urban fabric classification tasks, which serves as a baseline to compare different models. Then, we present a modified model, named “MMSSModel”, aiming to achieve better performance, by appending an auxiliary network for identifying the most appropriate combinations of the multiple spatial scales in line with the classification task.

4.1 Baseline model

CNN models developed for conventional image classification tasks, like LeNet, AlexNet, VGG16, etc., first extract features from the input image through a series of convolutional layers, pooling layers and normalization layers, and then transform and flatten the tensor to a dense vector, before processing by another series of fully connected layers followed by a softmax function to a value in the [0,1] range, which represents the probability that the input image belongs to the category rather than others [12, 14, 19]. Our baseline model is developed following this classic pipeline and is equipped with the ability to handle a set of inputs representing multiple spatial scales by adopting a straight-forward approach: directly concatenate the dense vectors transformed from multiple spatial

scale inputs and feed the newly formed vector into the classifier to perform classification.

Figure 5: Network structure of the baseline model. Given three different scales of inputs, the baseline model uses three different CNNs to extract the features respectively, then just fuse them by concatenation across dimension and feed into three classifiers to get outputs for different tasks.

More specifically, as shown in Figure 5, the input of the baseline model is a set of three 128×128×3 data patches, which pictorially represent a piece of 3-dimensional urban fabric in multiple spatial scales (representing 128m × 128m, 256m × 256m and 512m × 512m area). After the three data patches are transformed through three identical and independent CNNs respectively (we named them E128, E256 and E512), the three output dense vectors (V128, V256 and V512) are concatenated to form a longer one (MSV), which stores all the information and features extracted from the three input data patches. Then, MSV is fed into three classifiers for the three classification tasks (predicting the city origin, urban fabric pattern type and construction period). In each classifier, MSV processes by a fully connected layer followed by a softmax function to a series of values in the [0,1] range, which represent the probability distribution of the input urban fabric over all classes.

4.2 Modified model (MMSSModel)

Unlike the baseline model, which does not have a clear selector for choosing dimensions, leading to a more ambiguous classification process, MMSSModel is designed to learn the selection of the most appropriate combinations of the multiple spatial scales in line with the input multi-spatial scale data sample and the classification task through supervised learning. Then, it performs classification regarding the intelligently selected best scale combination. Total seven possible spatial scale combinations are identified and seven classification paths, which are developed based on the combinations are shown in Figure 6.

As shown in Figure 6, three “path selector” modules, named “PS_T1”, “PS_T2” and “PS_T3” are attached to the MSV and output three sets of path weight for the three classification tasks. Each set of weight contains 7 values in the [0,1] range, which are the predicted coefficients adopted to weight the seven scale combinations.

Figure 6: Network structure of MMSSModel. Same as the baseline model, MMSSModel extracts image features from different scales respectively, then the model will additionally predict specific spatial attention for each task to reveal the potential scale preference for each task. Finally, like the baseline model, MMSSModel will also predict three outputs for different tasks.

With the weights and classification results from different paths, the weighted classification result for each task can be calculated as weighted sum. More details about our network can be found in appendix materials.

5 EXPERIMENTS AND RESULTS

In this section, we evaluate the performance of the two models proposed in this work (the baseline model and MMSSModel) on the multi-spatial scale urban fabric dataset for Amsterdam, Barcelona, Berlin and Prague. Additional three classification models, which only take single spatial scale data as input (128px by 128px representing $128\text{m} \times 128\text{m}$, $256\text{m} \times 256\text{m}$ and $512\text{m} \times 512\text{m}$ region, named 128SSM, 256SSM and 512SSM respectively), are also included into the comparison to demonstrate the effectiveness of jointly considering urban fabric features across multiple spatial scales in the classification tasks.

5.1 Dataset and evaluation metrics

To evaluate our urban fabric classification models, the proposed multi-spatial scale urban fabric dataset for Amsterdam, Barcelona, Berlin and Prague were divided into two parts, training set (samples extracted from 60% of land parcels in each city) and testing set (samples extracted from the remaining 40% of land parcels). Since the full training set, which only contains 405 “open block” samples, 1081 “medieval region” samples and 3744 “orthogonal grid” samples across the four cities, is not balance, we implemented

over-sampling technique to balance the training set before the experiments. Additional data samples in minority classes were extracted with increased sampling density, from the land parcels that were used to sample training data. After the additional data samples were supplemented to the full training set, we randomly sampled around 10000 data samples for each city from the full training set to form a balanced training set.

F1-scores are taken as the evaluation criteria in order to measure the performance of category recognition results in this work. The F1-score is defined as the harmonic mean of precision and recall, where precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. These three metrics are calculated as $F1\text{-score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$, $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$ and $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$, where TP, FP and FN are the number of true positive, false positives and false negatives respectively.

5.2 Implementation details

In the training stage, We used softmax cross-entropy loss to train the three tasks respectively, and used Adam [10] with the learning rate of 4×10^{-3} which decayed at each 4 epochs with the decay rate of 0.2. The model was trained for 11 epochs totally. In addition, Dropout [20] was applied with ratio of 0.1 for Linear layers. Other settings were set by default.

	Category	128SSM	256SSM	512SSM	Baseline	MMSSModel
task1	Amsterdam	0.9273	0.9272	0.9567	0.9364	0.9533
	Barcelona	0.9460	0.9591	0.9826	0.9604	0.9797
	Berlin	0.9387	0.9428	0.9623	0.9593	0.9595
	Prague	0.8895	0.9088	0.9405	0.9088	0.9238
task2	Linear development	0.5795	0.6552	0.6961	0.7006	0.6971
	Open block	0.6240	0.7098	0.7430	0.7368	0.7452
	Gated compound	0.7338	0.7900	0.8155	0.8205	0.8131
	Medieval region	0.5883	0.6746	0.5974	0.6172	0.6663
	Irregular grid	0.6439	0.7288	0.7708	0.7683	0.7685
	Orthogonal grid	0.6133	0.6957	0.7394	0.7194	0.7258
task3	Before 1850	0.6090	0.7140	0.7122	0.6879	0.7431
	1850 to 1900	0.6746	0.7393	0.7869	0.7549	0.7873
	1900 to 1950	0.5485	0.6083	0.6436	0.6569	0.6504
	Post 1950	0.6557	0.7094	0.7299	0.7237	0.7301

Table 1: Performance of the models on the testing dataset on the three classification tasks, as measured by the F1-Score

5.3 Classification results

In the experiment, we ran each model three times and reported the average results. Figure 7 presents the average F1-scores of the baseline model and MMSSModel on the three tasks: predicting (1) the city origin, (2) urban fabric pattern type and (3) construction period. We noted that MMSSModel outperformed the baseline model on all tasks in terms of average F1-score.

Figure 7: Comparing different models based on average F1-Scores for the three classification tasks

To better compare different models and demonstrate the benefits of jointly considering urban fabric across multiple spatial scales, we present the F1-scores of the baseline model, MMSSModel together with 128SSM, 256SSM and 512SSM across all observations and on all tasks in Table 1.

As shown in Table 1, among 128SSM, 256SSM and 512SSM, 512SSM achieved the highest average F1-scores on all tasks and provided better results in most classes. However, smaller scale and more detailed information are required when observing the urban fabric patches in “medieval region” and “before 1850” class. 256SSM outperformed 512SSM on class of “medieval region” and “before 1850” by 0.0772 and 0.0018 respectively. This demonstrated that by sticking on single spatial scale, models could not always achieve the better performance on all class and all tasks. If we simply choose the single spatial scale model, which achieved the highest average F1-score (512SSM in this case) to perform classification across cities, the worst classification results may occur on some classes (0.5974 on “medieval region” in this case).

By jointly considering urban fabric in multiple spatial scales, our baseline model mitigates the problem and guarantees that the F1-scores on all classes obtained by the baseline model are higher than the lowest ones obtained by the single spatial scale models. At the same time, we noticed that the overall performance of our

Predicted Ground Truth	Amsterdam	Barcelona	Berlin	Prague	Precision	Recall	F1-score
Amsterdam	6440	63	110	217	96.44%	94.29%	0.9535
Barcelona	15	7119	35	72	97.65%	98.32%	0.9798
Berlin	146	35	7564	169	96.36%	95.58%	0.9597
Prague	77	73	141	4545	90.85%	93.98%	0.9239

Table 2: Confusion matrix of all classes for Task 1: predicting city origin, and precision, recall, F-Score per class

baseline model was worse than that of 512SSM, in terms of average F1-score. The reason is: in our baseline model, all information extracted across all spatial scales, which may contain noise and fault information from inappropriate spatial scales, are fed into the model and force the model to consider all of them when performing classification. As the model cannot distinguish the fault information from the useful ones, the information from inappropriate scales are treated equally by the model. This drags down the F1-scores of our baseline model and makes it unable to perform as good as the best single spatial scale classification models on all classes. To address this challenge, MMSSModel attached “path selector” modules to select the most appropriate combination of the multiple spatial scales in line with the input multi-spatial scale data sample and the classification task. This equips the model with the ability to ignore the fault information from inappropriate spatial scales. The average F1-scores of MMSSModel are higher than other models included in comparison on the two tasks: predicting urban fabric pattern type and construction period (shown in Figure 7). As shown in Table 1, MMSSModel achieved superior results on partial classes (4 out of 14 classes), in terms of F1-scores, when compared to the other models. Although there are exceptions, the differences in F1-scores are limited within 0.03.

Table 2, Table 3 and Table 4 show the confusion matrixes of the experiment on the three classification tasks and the precision, recall and F1-score on each class obtained by MMSSModel. In the confusion matrixes, the rows correspond to the ground truth classes and the columns correspond to the predicted classes. The table 3 reveals that urban fabrics in “medieval region” class are difficult to be distinguished from the ones in “linear development” class and the “orthogonal grid” patterns are easy to be confused with “irregular grid” patterns. These are consistent with the findings stated in *The City Shaped*, where the author assembled the medieval pattern and linear development pattern as “organic” pattern while grouping the irregular grid pattern and orthogonal grid pattern as “grid” [11].

6 FURTHER APPLICATIONS IN CITY MAPPING

Given the proof of the effectiveness of MMSSModel on urban fabric classification tasks, MMSSModel demonstrated its significant potential in city mapping, where we use the pre-trained MMSSModel to perform urban fabric classification across a city excluded in the training dataset. By mapping the predicted urban fabric type and construction period across a new city, the spatial distribution of different urban forms and the process of urban expansion can be intuitively presented. In practice, these help practitioners quickly understanding the spatial-temporal evolution of urban fabric. To achieve this, the generalization capability of MMSSModel is the key.

Predicted Ground Truth	Linear development	Open block	Gated compound	Medieval region	Irregular grid	Orthogonal grid	Precision	Recall	F1-score
Linear development	4824	47	1373	79	896	116	74.19%	65.77%	0.6973
Open block	67	200	6	0	5	2	78.43%	71.43%	0.7477
Gated compound	824	6	7435	11	331	82	77.46%	85.58%	0.8131
Medieval region	155	0	135	450	49	2	80.65%	56.89%	0.6672
Irregular grid	585	0	574	17	5707	339	74.81%	79.02%	0.7686
Orthogonal grid	47	2	75	1	641	1736	76.24%	69.38%	0.7265

Table 3: Confusion matrix of all classes for Task 2: predicting urban fabric type, and precision, recall, F-Score per class

Predicted Ground Truth	Before 1850	1850 to 1900	1900 to 1950	Post 1950	Precision	Recall	F1-score
Before 1850	1700	341	18	165	72.40%	76.44%	0.7437
1850 to 1900	366	6914	506	691	76.10%	81.56%	0.7874
1900 to 1950	103	1006	4347	1766	70.77%	60.19%	0.6506
Post 1950	179	824	1271	6624	71.64%	74.44%	0.7302

Table 4: Confusion matrix of all classes for Task 3: predicting construction period, and precision, recall, F-Score per class

Figure 8: Comparing different models based on average F1-Score for two classification tasks: predicting urban fabric type and construction period

To test the generalization capability of our proposed urban fabric classification model and illustrate the process of using our model in city mapping, we used the data samples from Amsterdam, Barcelona and Berlin to train MMSSModel and test it on the data samples extracted from Prague. All data samples are sampled from the proposed multi-spatial scale urban fabric dataset. As shown in Figure 8, the average F1-scores of MMSSModel on this generalisation capability test were dropped significantly, compared to those achieved in previous experiment, from 0.74 to 0.38 and from 0.73 to 0.34 on the two classification tasks: predicting urban fabric type and construction period respectively. However, the performance of the model was not dropped on all classes. The F1-scores on “linear development”, “gated compound”, and “irregular grid” maintained at similar levels as those in previous experiments. Only the ones on “medieval region” and “orthogonal grid” dropped significantly. These demonstrated that, in a new city, there are some unique characteristics in its urban fabric patterns which are never appeared in the training stage. Therefore, it is hard for our proposed model to handle such unknown traits without any reference. To address this challenge, we adopted the transfer learning approach [21]. First,

we trained MMSSModel using the samples extracted from Amsterdam, Barcelona and Berlin. Next, we used the trained model, named “ModelS1”, to process all the samples extracted from Prague and obtained the classification results with corresponding confidence levels. We then manually labelled 25% of data samples in low confidence levels and collected all these newly labelled data together with the data samples used to train ModelS1 to form a new training dataset. After that, a new model, named “ModelS2”, was trained on the new training dataset using the pre-trained ModelS1 as starting point. By adopting this approach, the average F1-scores of ModelS2 bounced back and reached 0.65 and 0.61 on the two classification tasks.

In Figure 9, we presented the urban fabric type mapping and construction period mapping for Prague, which were obtained using the trained ModelS2. The comparison between the predicted mappings and the ground truth ones demonstrated that MMSSModel can be utilized for new cities’ urban fabric type and urban process mappings, by only providing relatively small amount of additional labelled training data (provide only 25% labelled data and obtain similar performance as the ones trained using 60% labelled data). Moreover, with the proposed urban fabric dataset being expanded by including more urban fabric samples extracted from other major cities around the world, the ModelS1 can be expected to perform better and better, because more and more urban fabric patterns have been seen through the training stage. Thus, the number of additional labelled training samples required to train the ModelS2 will be reduced further.

7 CONCLUSION

Understanding the typology of cites and their evolution paths is a fundamental topic in urban planning, design and geography. Generations of scholars and practitioners aspire to grapple with the increasing variety and complexity of urban morphology. Currently, the spatial structure and fabric of cities are classified through intensive manual work by senior professionals with decades of experience, and even then, inconsistencies and errors occur.

In this work, we present a novel CNN based model that is adapted to classify urban fabric and demonstrate its efficacy using a purpose-built urban fabric data set. The data set integrates high resolution 3-dimentional built form data and hierarchical road network covering four European cities, with generic urban fabric patterns and broad development periods labelled out for training and validation purposes. The classification model presented in this paper is developed based on CNN and is capable of predicting urban fabric’s city origin, pattern type and construction period through a supervised multi-task learning process. The novelty of the algorithm is

Figure 9: Comparing predicted city mapping with ground truth ones. In urban fabric type mappings, samples in different pattern classes are represented by points in different colors followed the color code defined in Figure 3. In construction period mappings, the color code used is: blue-“before 1850”, green-“1850 to 1900”, red-“1900 to 1950” and yellow-“post 1950”

characterised by jointly considering urban fabric features across multiple spatial scales. It addresses a distinct challenge in urban fabric classification problem – the spatial scale of generic urban fabric features can vary significantly across areas of different development periods in one city and also across cities. The experiments demonstrate that the proposed multi-scale approach can capture a wide range of urban fabric features across cities and development periods, and the effectiveness can be improved by attaching an auxiliary network for identifying the most appropriate combinations of the multiple spatial scales in line with the classification task. Further, we test the generalization capability of the model and find that by providing a small number of additional labelled training samples, the model can be utilized to perform urban fabric classification for new cities. This suggests the promise that our model can be used in urban fabric mapping. In the future, we plan to expand the proposed multi-spatial scale urban fabric dataset and test semi-supervised and even “few-shot” classification approaches for urban fabric analysis [6, 8].

REFERENCES

- [1] M Berghauser Pont and P Haupt. 2007. The Spacemate: Density and the typomorphology of the urban fabric. *Urbanism laboratory for cities and regions: progress of research issues in urbanism* (2007).
- [2] Erwan Bocher, Gwendall Petit, Jérémy Bernard, and Sylvain Palominos. 2018. A geoprocessing framework to compute urban indicators: The MAPUCE tools chain. *Urban climate* 24 (2018), 153–174.
- [3] Christina Chatzipoulka and Marialena Nikolopoulou. 2018. Urban geometry, SVF and insolation of open spaces: London and Paris. *Building Research & Information* 46, 8 (2018), 881–898.
- [4] Jia Dong, Li Li, and Dongqing Han. 2019. New Quantitative Approach for the Morphological Similarity Analysis of Urban Fabrics Based on a Convolutional Autoencoder. *IEEE Access* 7 (2019), 138162–138174.
- [5] Rohinton Emmanuel and Koen Steemers. 2018. Connecting the realms of urban form, density and microclimate.
- [6] Li Fei-Fei, Rob Fergus, and Pietro Perona. 2006. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence* 28, 4 (2006), 594–611.
- [7] Lawrence Frank, Mark Bradley, Sarah Kavage, James Chapman, and T Keith Lawton. 2008. Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation* 35, 1 (2008), 37–54.
- [8] Yves Grandvalet and Yoshua Bengio. 2005. Semi-supervised learning by entropy minimization. In *Advances in neural information processing systems*. 529–536.
- [9] Martin Herold, Helen Couclelis, and Keith C Clarke. 2005. The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, environment and urban systems* 29, 4 (2005), 369–399.
- [10] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [11] Spiro Kostof. 1991. *The city shaped: Urban patterns and meanings through history*.
- [12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
- [13] Sylvia Lavin. 1992. *Quatremere de Quincy*. MIT press.
- [14] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.
- [15] Kevin Lynch. 1960. *The image of the city*. Vol. 11. MIT press.
- [16] Vahid Moosavi. 2017. Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet. *arXiv preprint arXiv:1709.02939* (2017).
- [17] Kin Ho Poon, JH Kämpf, SER Tay, NH Wong, and TG Reindl. 2020. Parametric study of URBAN morphology on building solar energy potential in Singapore context. *Urban Climate* 33 (2020), 100624.
- [18] Philipp Rode, Christian Keim, Guido Robazza, Pablo Viejo, and James Schofield. 2014. Cities and energy: urban morphology and residential heat-energy demand. *Environment and Planning B: Planning and Design* 41, 1 (2014), 138–162.
- [19] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [20] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.
- [21] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. 2018. A survey on deep transfer learning. In *International conference on artificial neural networks*. Springer, 270–279.
- [22] Hao Wu, Zhen Li, Keith C Clarke, Wenzhong Shi, Linchuan Fang, Anqi Lin, and Jie Zhou. 2019. Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science* 33, 5 (2019), 1040–1061.

APPENDIX

In this section, we will give a table to introduce the details of our model structure. Firstly, the inputs of our model are three scales of data with the dimension of $3 \times 128 \times 128$. As shown in this table, I_{128}, I_{256} and I_{512} in (1) represent the input data of three scales, f_{128}, f_{256} and f_{512} in (15) represent the features of three scales, w_1 -

w_3 represent the path selection weights for each classification task, p_1 - p_3 are the predictions of each path which is used to calculate the final weighted prediction p_{1f} - p_{3f} for each classification task. Here, the Index (1)-(14) are the CNN layers that are used to extract the features of inputs. And Index (15)-(28) are our designed layers for path selection and multi-prediction.

Index	Input	Operations	Output
(1)	$(3 \times 128 \times 128) \times 3: (I_{128}, I_{256}, I_{512})$	Convolution, Batch normalization, Relu activation	$(32 \times 128 \times 128) \times 3$
(2)	$(32 \times 128 \times 128) \times 3$	Max-pooling, Batch normalization, Relu activation	$(32 \times 64 \times 64) \times 3$
(3)-(14)	$(32 \times 64 \times 64) \times 3$	Repeat of (1) and (2)	$(1024 \times 2 \times 2) \times 3$
(15)	$(1024 \times 2 \times 2) \times 3$	Flatten layer, Linear layer, Relu activation	$(2048) \times 3: (f_{128}, f_{256}, f_{512})$
(16)	f_{128}, f_{256}	Concatenate, Linear layer, Relu activation	(2048): f_4
(17)	f_{256}, f_{512}	Concatenate, Linear layer, Relu activation	(2048): f_5
(18)	f_{128}, f_5	Concatenate, Linear layer, Relu activation	(2048): f_6
(19)	f_{128}, f_{256}, f_5	Concatenate, Linear layer, Relu activation	(2048): f_7
(20)	f_{128}, f_{256}, f_5	Concatenate, Linear layer, Softmax	(7): w_1
(21)	f_{128}, f_{256}, f_5	Concatenate, Linear layer, Softmax	(7): w_2
(22)	f_{128}, f_{256}, f_5	Concatenate, Linear layer, Softmax	(7): w_3
(23)	$f_{128}, f_{256}, f_5, f_4, f_5, f_6, f_7$	Linear layer, Relu activation	$(7 \times \text{numberofclass}_1): p_1$
(24)	$f_{128}, f_{256}, f_5, f_4, f_5, f_6, f_7$	Linear layer, Relu activation	$(7 \times \text{numberofclass}_2): p_2$
(25)	$f_{128}, f_{256}, f_5, f_4, f_5, f_6, f_7$	Linear layer, Relu activation	$(7 \times \text{numberofclass}_3): p_3$
(26)	w_1, p_1	Weighted summation	$\text{numberofclass}_1: p_{1f}$
(27)	w_1, p_1	Weighted summation	$\text{numberofclass}_2: p_{2f}$
(28)	w_1, p_1	Weighted summation	$\text{numberofclass}_3: p_{3f}$

Details of our network.