Article

Analysis of the Uniqueness and Similarity of City Landscapes Based on Deep Style Learning

Ling Zhao ¹ , Li Luo ¹, Bo Li ², Liyan Xu ³ , Jiawei Zhu ^{1,*}, Silu He ¹ and Haifeng Li ¹

¹ School of Geosciences and Info-Physics, Central South University, Changsha 410083, China; zhaoling@csu.edu.cn (L.Z.); luoli6680@csu.edu.cn (L.L.); hesilu@csu.edu.cn (S.H.); lihaifeng@csu.edu.cn (H.L.)

² School of Architecture and Art, Central South University, Changsha 410083, China; libo0910@csu.edu.cn

³ College of Architecture and Landscape Architecture, Peking University, Beijing 100871, China; xuliyan@pku.edu.cn

* Correspondence: jw_zhu@csu.edu.cn

Abstract: The city landscape is largely related to the design concept and aesthetics of planners. Influenced by globalization, planners and architects have borrowed from available designs, resulting in the “one city with a thousand faces” phenomenon. In order to create a unique urban landscape, they need to focus on local urban characteristics while learning new knowledge. Therefore, it is particularly important to explore the characteristics of cities’ landscapes. Previous researchers have studied them from different perspectives through social media data such as element types and feature maps. They only considered the content information of a image. However, social media images themselves have a “photographic cultural” character, which affects the city character. Therefore, we introduce this characteristic and propose a deep style learning for the city landscape method that can learn the global landscape features of cities from massive social media images encoded as vectors called city style features (CSFs). We find that CSFs can describe two landscape features: (1) intercity landscape features, which can quantitatively assess the similarity of intercity landscapes (we find that cities in close geographical proximity tend to have greater visual similarity to each other), and (2) intracity landscape features, which contain the inherent style characteristics of cities, and more fine-grained internal-city style characteristics can be obtained through cluster analysis. We validate the effectiveness of the above method on over four million Flickr social media images. The method proposed in this paper also provides a feasible approach for urban style analysis.

Keywords: city landscape; social media images; Flickr; style distance; visual similarity

Citation: Zhao, L.; Luo, L.; Li, B.; Xu, L.; Zhu, J.; He, S.; Li, H. Analysis of the Uniqueness and Similarity of City Landscapes Based on Deep Style Learning. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 734. <https://doi.org/10.3390/ijgi10110734>

Academic Editor: Wolfgang Kainz

Received: 2 September 2021

Accepted: 23 October 2021

Published: 29 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A city landscape (CL) is a visually perceivable characteristic of a city and is an important symbol of urban identity, regional culture, and urban charm and vitality. It is influenced by both physical and nonphysical environments, including city open space, building form, urban culture, and human activities. Cities are the concentrated manifestation of human civilization development, social changes, and lifestyle. Over the past thousands of years, a large number of cities with distinctive characteristics have been formed, such as Beijing with its ancient capital charm and Sydney as a “harbor city”. However, in the context of globalization, many cities have gradually lost their characteristics, and the problem of “one city with a thousand faces” has emerged. It is mainly manifested in two aspects [1]: (1) the dilution of cultural traditions, which causes the lack of perception and identity of urban places, and (2) in urban construction, planners and architects introduced designs that were available anywhere, lacking originality. Gradually, they always had a wide range of shared patterns and international thinking habits. The urban landscape is largely related to the design philosophy and aesthetics of planners. To create a city with character, they need not only to learn new knowledge but also to have a deep awareness and understanding

of the local landscape features. Therefore, it is important to explore the characteristics of the urban landscape for urban planning, urban design, and cultural communication. It can provide urban planners/designers with more original design concepts to improve the attractiveness and charm of cities, and it can make us know more about urban culture. In the past decade, the creation and planning of CLs has received more attention, but because cultural characteristics are difficult to measure and the degree of uniqueness or similarity between cities is not easy to judge, scientific quantitative methods and objective analysis techniques applicable to CL construction still need to be improved.

To depict and represent a CL, it is helpful to measure the degree of uniqueness and similarity between cities. Earlier studies mainly relied on questionnaires and interviews to explore CL characteristics [2–4], but it is difficult for these traditional methods to obtain a large number of research samples, thus affecting the objectivity and validity of any study. With the rapid development of various social media software (e.g., Flickr, Weibo, Instagram) and Web mapping services (e.g., Google, Tencent), the number of urban images has increased exponentially, covering every corner of a city. With new opportunities made available by these images [5], researchers have also gradually started to focus on the exploration and use of urban image data to study CLs. For example, Shalunts et al. explored different building facade window styles (Gothic, Baroque, and Romanesque) [6] and dome styles [7] based on Flickr images. Doersch et al. [8] explored the visual elements that can best represent the urban qualities of Paris with Google Street View images to understand which kinds of balconies or windows in Paris look most similar. In recent years, due to the rapid development of deep learning in the field of computer vision, convolutional neural networks (CNNs) with powerful learning and expression capabilities have made breakthroughs in tasks such as image classification [9–11], image scene recognition [12–14] and change detection [15,16], making it possible to accurately and rapidly mine richer information from massive social media images. Thus, using deep learning methods to deeply understand and explore CLs from the vast amount of image data generated by social media and online maps is a new research interest. Based on this development, researchers have started to study the appearance of cities in depth, such as identifying the architectural styles of Mexico [17], the architectural elements of specific periods and analyzing how functionally similar architectural elements change over time [18], and automatically identifying the age of buildings [19]. Deep learning was even used to simulate the human brain to perceive the city's surroundings [20–23] to explore what makes London look beautiful, quiet, and happy [24]. Many studies have shown that deep-learning-based models are indistinguishable from humans in their perceptual abilities and may be superior. In addition, researchers have explored city identity or elements from a large number of geotagged images [25–27], measured the similarity of urban scenes and objects, and discovered the uniqueness of a city [28]. Based on the above efforts, deep learning methods, particularly CNNs with the ability to extract excellent features and to master high-level task information for complex scenes, help to better capture CLs, which is beneficial for our research.

The intention of this study is to acquire CLs from a large number of images to measure the visual differences among cities. The visual differences are mainly influenced by the images themselves, which have a “photographic cultural” nature [29]. This property is also responsible for CL similarity, and, therefore, we consider introducing this property, which has not been considered in previous studies. Ref. [29] also referred to this property as a “style feature”, and related studies [30,31] indicated that the statistical information of the feature maps of convolutional neural networks can represent such features well. Inspired by this, we introduced the “style feature”. From the perspective of feature maps, there is a hierarchical nature of the features in the network; shallow layers record basic information such as color and texture, while deeper layers record more advanced information that is class-specific and can be utilized to recognize full objects [32]. To obtain more useful information, we consider using statistical information on deeper features. To achieve our goal, we gather 534,767 social media images from 10 cities and use the mean and

variance of the feature maps of a four-layer CNN to compose the city style feature (CSF) in this paper to discover city landscape characteristics. To quantitatively describe the differences in CL between cities, we define CL distance and measure the similarity and uniqueness of cities as a whole. In this paper, instead of setting specific criteria for CL identification, we use an unsupervised approach to discover CL types. In addition, since cities in different eras and regions are certain to have different landscape features, we analyze the more detailed landscape features of cities. Therefore, we propose a clustering analysis approach for fine-grained CL that constitutes the overall characteristics. Our contributions are as follows:

- (1) We propose a CL representation method based on deep style learning and encode city style as a vector. Additionally, to solve the imbalance problem of social media images and to allow the network to better learn the style features of cities, we assign different weights to each category.
- (2) We define CL distance using the CSF to analyze how different cities represent landscape similarities and differences. We find that cities in close geographical proximity tend to have greater visual similarity to each other.
- (3) To deeply understand the landscape characteristics of individual cities, we use a clustering method with the CSF as the embedding vector that can discover the fine-grained landscape features of cities in a more detailed way.

The rest of this paper is organized as follows. In Section 2, we discuss related work. In Section 3, we describe the source of the data set used for the experiments in this paper and the preprocessing of the data set and our method. In Section 4, we report the experimental results and analysis. Finally, in Section 5, we provide discussions and limitations. In Section 6, we summarize our findings and propose future research directions.

2. Related Work

2.1. Comparison between Social Media Imagery and Street-Level Imagery

Currently, the main sources of urban images are (1) social network platforms (e.g., Facebook, Weibo, Flickr, Twitter) and (2) map service platforms (e.g., Google, Baidu). We call images from (1) “Web imagery” that are taken and uploaded by users with different shooting angles and various objects but that provide an overall perception of a city. By contrast, the data from (2) that we generally call “street-level imagery” have a uniform shooting angle and a more uniform image sampling distribution, and the recoding contents are generally determined by the research objectives.

The two data sources have some differences and similarities. In Table 1, we compare Web imagery with street-level imagery. Both Web imagery and street-level imagery can cover every corner of a city, but Web imagery has a bias toward areas with characteristics of a city, has certain advantages in studying CLs, and can better discover different scenes with historical and cultural atmospheres in a city. Web imagery is mainly used for tourism [33] and urban areas of interest [34] and urban characteristic analyses. Street-level imagery is mainly used for predictive analysis [20] and urban safety analysis. Zhou et al. [25] used Web imagery to analyze the urban element types of seven cities and explored the similarities and differences between cities. Other scholars used Web imagery to evaluate the imagery characteristics of different cities in terms of the overall distribution structure and uniqueness of the cities [35]. Kita [36] used Google Street View images of houses to predict the risk of car accidents and proposed a risk prediction model. Saleses et al. [37] used Google Street View images to analyze the street safety of four cities. Based on the comparison of the two types of image sources, we chose to use social network imagery in our study.

Table 1. Comparison between Web imagery and street-level imagery.

	Web Imagery	Street-Level Imagery
Data Source	Facebook, Weibo, Flickr, Twitter, etc.	Google Map, Baidu Map
Shooting Angles	Nonuniform	Specific angles
Application	Analysis of tourist routes, analysis of city uniqueness, etc.	Traffic flow prediction analysis, urban safety analysis, etc.
Recoding Contents	various objects, including indoor and outdoor imagery	Generally outdoor imagery, e.g., buildings, bus stations, etc.
Motivation	Subjective	Objective

2.2. Computer Vision and Style

Style is an abstract concept that includes the artistic style, picture style, and fashion style and so on. Different definitions are given in different studies. In recent years, a great deal of work has been conducted to study styles in computer vision. On the application side, a realistic photograph is rendered into a nonreal image with artistic style, i.e., “style transfer” [31,38]. On the analysis side, researchers established and marked large data sets and classified style types. They discovered and analyzed similar or consistent visual styles using supervised, unsupervised, and visual consistency methods. Matzen et al. [39] made a large clothing data set annotated with 12 clothing attributes. They discovered multiple fashion combinations by clustering and performed a comparative analysis of Northern and Southern Hemisphere clothing. Redi et al. [29] analyzed the cultural styles of photographs using target detection methods and aesthetic computational tools to quantify the degree of similarity of photographs taken using a supervised classification approach. Shen et al. [40] proposed a visual consistency approach to find the same regions from artworks through cosine similarity.

In addition, learning and extracting style is a highly important module. Thus, many researchers have started to study the style features that are helpful for style classification. Karayev et al. [41] proposed a feature extraction method based on CNN. They demonstrated that their proposed method is more effective than traditional aesthetic feature methods with Flickr80K, Wikipaintings, and AVA style data sets. However, another work [17] mainly modified the CNN structure or combined multiple CNNs to improve the learning of style features.

In this paper, we propose a city style feature learning method. We use this method to discover the fine-grained style of individual cities.

3. Materials and Methods

3.1. Data Set and Study Area

The data set in this paper originates from the YFCC-100M (Yahoo Flickr Creative Commons 100 Million) data set [42], which contains the metadata of all videos and photos uploaded between 2004 and 2014, including download links, upload times, geolocations, user comments and machine tags, latitude and longitude, and 23 other dimensions of information, with more than 100 million data points. To explore the similarities and differences in the CL at different locations, we selected 10 cities located on four continents (Asia, Europe, North America, and Oceania). The richness of the sample facilitated our experimental analysis. Ref. [43] shows that there are more samples from the USA, Canada, China, and Australia. Taking into different cultures, urban landscapes, and economic factors into account, we have selected cities from the above countries that are part of the Global Cities. In addition, Tokyo and Paris are indispensable. Therefore, these cities are Beijing, Shanghai, Hong Kong, Tokyo, Toronto, New York, Montreal, Paris, London, and Sydney, with a total of 4,387,980 images collected.

3.2. Data Preprocessing

We are interested in the CL characteristics. YFCC-100M derives from crowdsourcing, and there exists a large number of images such as interiors, people, flowers, animals, airplanes, and sky images. These images may have some unpredictable effects on the

experimental results. Therefore, we consider all of these images as noisy samples in our study. In this paper, we design a two-stage denoising method.

1. First stage: indoor and outdoor images—automatic coarse rejection

Our research targets images that represent distinctive urban scenes that prioritize outdoor scenes. Therefore, we formulate the finding of outdoor scenes as an image binary classification task that aims to automatically reject noisy samples. In this paper, we use an indoor–outdoor biclassification model trained with the Place365 data set [25] to classify 10 city images. Finally, a total of 750,850 outdoor scene images are retained.

2. Second stage: outdoor noise image fine rejection

After the above processing, some nonrepresentative images, such as flowers, animals, airplanes, and skies, are still present. Thus, it is necessary to perform further filtering. Considering that flowers, animals, airplanes, and skies display obvious differences from outdoor urban landscape representative objects such as buildings and bridges, we use clustering to reject this noise. The specific steps are as follows:

- We train a classifier (ResNet50) with city names as categories. For each city, we randomly selected 5000 images as training samples. The training parameters and details can be found in Section 3.3.
- We directly use the features of the pooling layer as the input for clustering, and the number of clusters for each city is 30 (set based on our experiment).

We found that the majority of the noisy and non-noisy samples were well distinguished and clustered into their respective classes. As a result, we obtained 534,767 images, with the specific number distribution shown in Table 2.

Table 2. Total number of samples for each city.

City	Number	City	Number
Beijing	29,604	Shanghai	15,376
Hong Kong	37,724	Tokyo	86,044
New York	107,967	Sydney	23,108
Toronto	28,585	Montreal	11,148
Paris	73,487	London	121,724

3.3. Methods

This section focuses on the overall framework of our method, as shown in Figure 1. To learn the city style features used for various analyses, we use a convolutional neural network to automatically learn the rich internal feature hierarchy in the given training set, where we count the mean and variance of the feature maps in the fourth layer of the network and input the vector composed by them after connecting them as landscape features into the fully connected layer, as shown in Figure 1b. In this section, we also describe the training techniques used in the experiment (Section 3.3.4). To measure the similarity of landscape characteristics between cities (Section 3.3.2) and to discuss the fine-grained landscape characteristics of cities (Section 3.3.3), we also describe the corresponding methods in this section.

3.3.1. Style Learning and City Style Feature

(1) Deep Style Learning

City landscape can describe the global style feature of a city. Inspired by Sergey et al. [41], we model city landscape as the style feature learned by convolutional neural networks from massive images from a city. Sergey et al. [41] also shows that style features learned by convolutional neural networks outperform traditional manual features. Therefore, the approach in this paper is based on a CNN, more specifically, on the ResNet-50 neural network architecture, which is composed of two averaging pooling layers and four residual network blocks. The shallow layers of the network represent low-level features (e.g., edges

and textures), while the deep layers represent abstract features (target objects or semantics). However, we are more concerned with the target object.

Figure 1. Overall framework. B is short for batch size.

Since image styles are diverse and abstract, we need a general method that can represent arbitrary image style. Ref. [31] shown that the mean and standard deviation of each channel of the feature map extracted by a convolutional neural network can represent the appearance of an arbitrary image. As seen in Figure 2b, it meets our requirements. Based on this, we implement the style learning. Figure 2a illustrates the process of style learning. We trained a classifier with classification labels of cities name, which feeds into city images and outputs the predicted probability values. In addition, we counted the mean and variance of l th layer feature maps. These two statistics are concatenated with $dim = 1$ to form the city style feature (CSF), which is used as the input to the fully connected layer. The network is iteratively updated through a forward and backward propagation process to achieve city style learning. Figure 2a(1) and Figure 2a(2) show the distribution of the statistical values and the CSF, respectively, where the number of channels are the horizontal axis and their values are the vertical axis.

Figure 2. Deep style learning process and examples of city style feature.

(2) City Style Feature

CSF is mainly used to represent the global feature of an image. Next, we will give its definition. Given an input image $x_0 \in R^{W_0 \times H_0 \times 3}$, where W_0 and H_0 represent the image width and length, a convolutional neural network maps x_0 into a set of feature maps $\{F^l(x_0)\}_{l=1}^L$, where $F^l: R^{W_0 \times H_0 \times 3} \rightarrow R^{W_l \times H_l \times N_l}$ is the mapping from the image to the l th layer tensor activations, where the spatial dimension of N_l channels is $W_l \times H_l$. In this paper,

we reshape the activation tensor $F^l(x_0)$ into a matrix $F^l(x_0) \in R^{N_l \times M_l}$, where $M_l = W_l H_l$. Therefore, based on 4.1 (1), the city style feature in the l th layer can be expressed as:

$$CSF = \text{concat}(\mu(F_l), \delta(F_l)) \quad (1)$$

where μ and δ denote the mean and variance, respectively, as $\mu(F_l) \in R^{N_l \times 1}$, $\delta(F_l) \in R^{N_l \times 1}$. In this paper, we set the $l = 4$, and the CSF has 4096 dimensions features where $N_l = 2048$.

3.3.2. Intercity Landscape

To enable the network to better learn the landscape features among cities, we train a network with city names as categories. According to Equation (1), we calculate the layer 4 $CSFs$ of the convolutional neural network as the landscape feature among cities.

To quantitatively describe the similarity among cities, we indirectly calculate the landscape distance using the CSF . The single-target (e.g., building) similarity metric is simple. However, multiobjective similarity metrics are complex, such as similarity metrics between cities. In previous studies [28,29,44], confusion matrices are often used to solve multiobjective metric problems. This method has the advantage of simple calculation and no consideration of the type and number of targets. Based on this, this metric is used in this paper to achieve the calculation of landscape distance.

Definition of the landscape distance (LD): If two cities are similar, their samples are more likely to be misclassified into each other. Thus, LD can be calculated by the misclassification rate of two cities. Suppose city C_i has S_i samples misclassified into city C_j , and city C_j has S_j samples misclassified into city C_i ; then, LD between C_i and C_j can be expressed as:

$$LD_{i,j} = \text{Norm}(S_i) + \text{Norm}(S_j) \quad (2)$$

where $i \neq j$. Norm is a normalization operation to ensure uniformity of magnitude.

3.3.3. Intracity Landscape

The composition of cities is complex and diverse, and each city is composed of different elements. We would like to further explore the CL to analyze the style type, such as “what are the main components of Beijing’s landscape?”, which we do in Section 4.3. During the experiment, we find that the CSF can describe not only the intercity landscape but also the intracity landscape. In previous studies, the landscape type was generally explored and analyzed with landscape labels in a supervised manner, but the landscape type was difficult to define. To achieve our goal, we use clustering to identify similar visual patterns in the landscape’s embedding space.

We calculate the intercity landscape in the same way as in Section 3.3.1, except that the target becomes a single city. To find the landscape type, we run the clustering algorithm on a subset (60% in total) of full samples from a single city for efficiency. We compute $2N_l$ -D CNN feature vectors that contain much-repeated information. To reduce the information redundancy, we use PCA to project these vectors onto the subject principal components that retain 90% of the variance (in our case, 259 dimensions). To cluster these vectors, we use a Gaussian mixture model (GMM), which is more flexible in handling cluster groups of multiple shapes. By giving clustering components, each image is allocated to the component with the maximum probability.

To avoid arbitrarily assigned clustering numbers that lead to overfitting, we use the criterion provided by the scikit-learn library for determining the number of constituents, namely the Akaike information criterion (AIC).

3.3.4. Training Tricks

(1) Pre-Training Techniques

Convolutional neural networks can learn a sufficient number of features with a large number of training samples, but they are prone to overfitting and long training times. To

avoid these problems, transfer learning is introduced in this paper. It can help to solve existing problems by using existing knowledge and improve the robustness of the model. Transfer learning has been widely used in fields such as natural language processing [44], natural image classification [45–47], and target detection [48–50].

The ImageNet pretraining model contains features of 1000 classes, and is a better choice for our work. In this paper, the training weights of the base layer of the ImageNet pretraining model are fixed, while the weights of the fully connected layer are fine-tuned.

(2) Imbalanced sample

The experimental data set in this paper has a sample imbalance problem (as shown in Table 2). Inspired by [51], we introduce a penalty on the city samples so that the penalty is larger for cities with more samples, and vice versa. In this paper, we solve the imbalance of samples by setting the value of α . It is expressed as follows:

$$\text{loss}(p) = \begin{cases} -\alpha(1-\alpha) \log p, & y = 1 \\ -(1-\alpha)p \log(1-p), & y = 0 \end{cases} \quad (3)$$

where α is the weight of each city. A larger number of samples corresponds to a smaller weight. $\alpha_i = \text{Num}_{\min} / \text{Num}_i$, ($i = 1, 2, \dots, N$). $\text{Num}_{\min}$ is the minimum number of samples. Num_i is the number of samples of the i th city. $p \in [0, 1]$ is the estimated probability of the model for the class labeled $y = 1$.

3.3.5. Training Details

We train the CNN as follows. The data set is split into training set, validation set, and test set with the proportion of 6:2:2. The validation set is mainly applied to adjust the parameters during the training of the model to determine when to stop training. The batch size is 1024. We use the forward and backward propagation of CNN to calculate the parameter gradient of the loss function. Because the original images vary in size, the image scale input to the network for training is 256. We update the parameters of the network using stochastic gradient descent with $\text{momentum} = 0.9$, learning rate = 0.001, and weight decay = 10^{-4} . We train the CNN for 800 iterations using a cosine annealing learning rate decay strategy. Finally, we achieve an average accuracy of 49.9%. We make predictions on the test set and use the confusion matrix to present the prediction results (Figure 3).

Figure 3. Normalized confusion matrix of classification results.

4. Results

4.1. Visual Analysis of City Landscape

(1) Unique visual style analysis

The confusion matrix obtained from the image classification task based on *CSF* shows the visual connections between cities, from which we can analyze the visual similarities and uniqueness of the landscape. The numbers in Figure 3 represent the normalized values of images from the locations showed in the column labels that are classified as originating from the cities shown in the row labels. If a city has a higher percentage of correctly classified images, it has a more unique visual landscape pattern. Instead, the higher the rate of misclassification between two cities, the higher the visual similarity of their landscapes. For example, Beijing was correctly classified with the highest percentage of 0.68 but was easily misclassified as Shanghai (0.06) and Tokyo (0.05).

The confusion matrix shows that Beijing (0.68), Sydney (0.50), Paris (0.49), Hong Kong (0.47), and Shanghai (0.46) are the five cities with the highest probability of correct prediction, indicating that these cities are more visually distinctive than other cities. On the other hand, five cities, namely Toronto (0.38), London (0.30), Tokyo (0.42), Montreal (0.28), and New York (0.28), are prone to incorrect predictions and are more likely to be confused and visually less unique than other cities.

To clearly understand the visual landscape of each city, we show some samples of each city with a high confidence of correct classification (shown in Figure 4). We found that the close-up and remote views show the way people record the city, and show that people view the city from different perspectives; for instance, people like to observe Beijing from close up, while they like to record Hong Kong from a distance. In addition, we also found that historical sites, landmarks, and unique urban landscapes are scenes with urban uniqueness in these cities. The historical buildings, the Forbidden City and Temple of Heaven, as well as landmarks such as Tiananmen Square, are the scenes with visual uniqueness in Beijing. Hong Kong mainly has a unique view of Victoria Harbor at night and from a distance. London's Tower Bridge and Big Ben are the factors that make it different from other cities. Montreal's Notre Dame Cathedral landmark makes Montreal more unique visually. In addition, landmarks such as the Brooklyn Bridge and the Empire State Building in New York, the Eiffel Tower and the Arc de Triomphe in Paris, the Oriental Pearl and the Yangtze River in Shanghai, the Sydney Opera House and the Sydney Bridge in Sydney, the Tokyo Tower, Sky Tree, Asakusa Temple, and other historical buildings in Tokyo, and the CN Tower and other landmarks in Toronto are the elements that make a city visually unique and are the scenes that present a visual difference from other cities. It is important to note that cars appear in both Montreal and Toronto. Montreal has racing that is related to its culture because racing events are annually held at the Montreal track, while for Toronto, the bus is a major distinctive feature.

(2) Similarity measures analysis

To quantitatively describe the visual similarity between cities, we calculate the visual landscape distance between two cities using landscape distance and obtain the visual landscape similarity matrix (as shown in Figure 5).

From the similarity matrix, we find that London is most similar to Paris (0.24), followed by Shanghai–Hong Kong (0.20), Hong Kong–Tokyo (0.20), Toronto–Montreal (0.19), and Toronto–Tokyo (0.19). We show the 10 samples with high misclassification probability for each of these 5 cities in Figure 6.

Figure 4. Samples of each city with unique visual landscape.

Figure 5. Normalized landscape similarity matrix.

(3) Visual Similarity Analysis

According to our experimental results, we found that there are similarities between different cities. Figure 6 shows the scenes with visual similarity between two cities that are easily misclassified:

- London–Paris (0.24): The architectural styles of the two cities are similar. We found that both cities have pointy roofs, which may be one of the reasons for their similarity. To verify this idea, we mapped the heat map of layer 4 using class activation mapping (CAM), a technique for visualizing which parts of an input image play an important role in model decisions. (See [52] for details of the CAM implementation.) As illustrated in Figure 7a, the roof is indeed one of the important points of interest for the model. In addition to this, arches, patterns of buildings, and window styles are also of interest to the model, indicating that these are the visual factors that make the two cities visually similar.
- Shanghai–Hong Kong (0.20): The panoramic view (distant view) is more similar. Similarly, we have analyzed the reasons for the visual similarities between Shanghai and Hong Kong. From Figure 7b, we found that the tall buildings and ports are the main factors making them visually similar. In fact, they have many similarities in geographical location and urban culture. Geographically, they are all coastal cities and both use ports to develop their economies. In terms of urban culture, both are open to the outside world. This has given them both a modern atmosphere.
- Hong Kong–Tokyo (0.20): We can see that the streets of Hong Kong and Tokyo are similar. At the same time, we find that in addition to the similar high-rise buildings, some ancient buildings in Hong Kong are quite similar to the style of temples in Tokyo.
- Toronto–Montreal (0.19): The architectural design and style are similar.
- Toronto–Tokyo (0.19): Mainly, the buildings are misclassified, and the window textures of modern buildings are more regular overall. In addition, we observe that the overall tone, camera angle, and texture are similar for the two cities.

Figure 6. Top five pairs of city samples with the greatest landscape distance.

Figure 7. Analysis of the reasons for the (a) London–Paris and (b) Shanghai–Hong Kong similarity. The first column (Image) represents the source image, and the second column (CAM) represents the class activation map of layer 4. In the CAM, the more red the color, the greater the effect.

4.2. Landscape Distance and Geographical Distance

To further investigate the relationship between urban landscape similarity and geographic location, we show the landscape distance and spatial location. Figure 8 shows the relationship between urban landscape similarity and the geographic location of cities. The cities connected by grey lines are the pairs of cities with high similarity scores, and the number above is the distance between them; the red circle identifies the geographic location of the city, and its color shade is positively related to the proportion of the city's image being correctly classified, i.e., positively related to the diagonal value in the confusion matrix.

In Figure 8, it is easy to see that cities in close geographical proximity tend to have greater visual similarity to each other, such as Montreal and Toronto, London and Paris, and Shanghai and Hong Kong. However, at the same time, we can also see that Montreal–New York and New York–Toronto are also geographically close, but the similarity is relatively low. As the visual characteristics of cities are largely influenced by culture, history, climate, and geography, their similarity is not necessarily high despite geographical proximity. Geography is not the only influencing factor.

Figure 8. Visual similarity of cities in relation to their geographical location.

4.3. Fine-Grained Intracity Landscape Feature

The city landscape features proposed in this paper can characterize not only the intercity landscape but also the intracity landscape, which is what we mainly explore in this section. Since there are more cities, it is too tedious to analyze them one by one. According to the similarity analysis in 5.1, Beijing has the most individual characteristics compared with other cities; thus, we only analyze Beijing. Using the method in Section 3.3.3 to obtain the clustering results, we selected some samples from clustering centers that were close to and representative of the clustering center (Figure 9). In Figure 9, we can clearly observe the characteristics of Beijing. For further explanation, we roughly divide the results into five main categories: Beijing's ancient buildings (a), target objects (b), modern landmarks (c), some unique landscapes (d), and Beijing at night (e).

Figure 9. Beijing fine-grained clustering samples display. Beijing can be divided into roughly 5 main categories: ancient buildings (a), target objects (b), modern landmarks (c), some unique landscapes (d), and Beijing at night (e).

The design of ancient buildings in Beijing is special, generally symmetrical from left to right, with the middle part slightly higher, mainly to reflect the supreme authority of the ancient Chinese emperors. In addition, the color of the walls is generally brick red. At the same time, the roofs of ancient buildings and the front of the houses are often accompanied by some auspicious patterns, such as dragons and lions, and some incense burners are placed in front of the houses (Figure 9b first three rows). Among the modern buildings, the CCTV headquarters building, the Great Hall of the People, and some high-rise buildings have attracted attention due to their design and role, thus forming one of the characteristics of Beijing (Figure 9c first five rows). The Great Wall, Tongyun Bridge, and 17-hole Bridge characterize the beautiful scenery of Beijing. Historical human interest and beautiful scenery do not always leave a deep impression. The closer one is to life, the deeper your feelings will be. The last two lines of Figure 9c show the hutongs of old Beijing. Hutongs are the best way to show older people's feelings about old Beijing and the life of old Beijing. The formation of Beijing's landscape is closely related to China's history, culture, and development.

5. Discussion

Most of the existing research objects are relatively single, while the landscape of a city should be diverse. Urban big data are being generated at an unprecedented rate, which creates new opportunities for studying the urban landscape from a multiobjective perspective. However, it is a challenge to characterize the urban landscape and quantify it among cities because cultural characteristics are difficult to measure. The excellent learning ability of convolutional neural networks helps to characterize urban landscapes. In addition, we believe that images are the carriers of visual information of cities and have their "style" property. Therefore, in this study, we propose a deep-style-learning-based urban landscape representation method to make a comprehensive quantitative comparison of the urban landscape. In this study, we show how to characterize the urban landscape and explore the visual differences in the landscape of 10 cities, and further analyze the composition of the landscape of individual cities. We found that historic buildings, vehicles, and unique landmarks are the scenes that make cities unique. Streets and spatial structures are the factors that make cities similar in appearance. There is often a greater visual similarity between cities that are geographically close to each other. This work may have the following implications. First, in urban planning, planners need to have an overall understanding of a city to know which places, scenes, or elements make the city unique, and to maintain and continue the characteristics of the city based on respecting the original landscape. Second, it is beneficial for people to explore and interpret the local culture. For example, Montreal is a racing city with racing history and tradition, and Shanghai and Hong Kong have more developed economies. Last but not least, understanding the landscape of a city is not only to focus on landmarks/historical buildings but also on corners that are characteristic in themselves but not often seen, and through fine-grained analysis, it is beneficial to find such hidden places, as in the last two rows of Figure 9c. For tourists, it can be used as a reference for hitting places when traveling.

Limitation. Although we achieved a relatively good result, there are some limitations to our method. YFCC100M is a shared metadata data set collected by Yahoo on the Flickr social platform. However, the data on Flickr are uploaded by users with different preferences at different times and places, so there will be some bias in the sample data, mainly in two aspects: (1) **Recording content.** The Web imagery is the users' perceptions and records of a city. The users like to record places with characteristics in a city and places they are interested in, so the images used in this paper will be biased in terms of content. However, we can more easily find places with characteristics in a city to analyze the appearance of a city. (2) **Geography.** There are two types of geographic locations. One is the location manually edited by the user, and the other is the autolocation position according to the shooting-equipped positioning system. For the first type, the positioning will be inaccurate.

Theoretically, our method proposed in this paper is general, but the results obtained are to some extent biased due to the deviations in the data itself. Therefore, our experimental results are only analyzed for the data used in this paper.

6. Conclusions

A city is formed through tens of thousands of years of history and is a concentrated expression of human civilization construction. Historic buildings, unique natural scenery, streets, and landmarks [53,54] are part of a city's landscape. In this paper, we propose a deep-style-learning-based urban landscape representation method that can handle multiple scenes or multiple targets. We call the city style features learned using our method as *CSF*. Experiments show that a *CSF* can not only distinguish the overall style between different cities but also further distinguish the local style within a city. Furthermore, we analyze 10 cities around the world with respect to two main aspects: (1) CL distance is defined using the *CSF* that was used to analyze how different cities are characterized as similar and are similar in style. (2) To deeply understand the CL characteristics of individual cities, we use *CSF* as an embedding vector for clustering analysis to discover the fine-grained CL in more detail. In addition, we found that although the two cities are geographically similar, their similarity may not be high.

Future work. The city landscape is contemporary and regional in nature. The urban values of each city are influenced by the prevailing culture at that time. Therefore, in our future work, we will analyze city similarity from two aspects—temporal and spatial.

- **Temporal.** We can analyze the similarity of several cities in different periods with the Flickr data timestamp.
- **Spatial.** Flickr metadata includes geographic coordinates and many comments about life and urban areas. These comments can be used as auxiliary information for the type of region or CL.

Author Contributions: Conceptualization, Ling Zhao and Bo Li; formal analysis, Liyan Xu and Jiawei Zhu; methodology, Ling Zhao, Li Luo, Bo Li and Haifeng Li; funding acquisition, Li Luo and Haifeng Li; supervision, Ling Zhao; visualization, Liyan Xu and Silu He; writing—original draft, Ling Zhao and Li Luo; writing—review and editing, Ling Zhao and Li Luo. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded supported by the Fundamental Research Funds for the Central Universities of Central South University under Grant 2021zzts0859 and by the National Natural Science Foundation of China under Grant 42171458, and by the Natural Science Foundation of Hunan Province under Grant 2021JJ30818.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available for download at <http://www.multimediacommons.org/>.

Conflicts of Interest: Not applicable.

References

1. Relph, E. *The Modern Urban Landscape*; Routledge: Oxfordshire, UK, 2016.
2. Milgram, S. A psychological map of New York City. *Am. Sci.* **1972**, *60*, 194–200.
3. Twigger-Ross, C.L.; Uzzell, D.L. Place and identity processes. *J. Environ. Psychol.* **1996**, *16*, 205–220. [[CrossRef](#)]
4. Paasi, A. Region and place: Regional identity in question. *Prog. Hum. Geogr.* **2003**, *27*, 475–485. [[CrossRef](#)]
5. Martí, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Social media data: Challenges, opportunities and limitations in urban studies. *Comput. Environ. Urban Syst.* **2019**, *74*, 161–174. [[CrossRef](#)]
6. Shalunts, G.; Haxhimusa, Y.; Sablatnig, R. Architectural style classification of building facade windows. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 26–28 September 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 280–289.

7. Shalunts, G.; Haxhimusa, Y.; Sablatnig, R. Architectural style classification of domes. In Proceedings of the International Symposium on Visual Computing, Rethymnon, Greece, 16–18 July 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 420–429.
8. Doersch, C.; Singh, S.; Gupta, A.; Sivic, J.; Efros, A. What makes paris look like paris? *ACM Trans. Graph.* **2012**, *31*. [\[CrossRef\]](#)
9. Sun, Y.; Xue, B.; Zhang, M.; Yen, G.G.; Lv, J. Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE Trans. Cybern.* **2020**, *50*, 3840–3854. [\[CrossRef\]](#)
10. Ma, B.; Li, X.; Xia, Y.; Zhang, Y. Autonomous deep learning: A genetic DCNN designer for image classification. *Neurocomputing* **2020**, *379*, 152–161. [\[CrossRef\]](#)
11. Zhu, Q.; Liao, C.; Hu, H.; Mei, X.; Li, H. MAP-Net: Multiple Attending Path Neural Network for Building Foot-print Extraction From Remote Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6169–6181. [\[CrossRef\]](#)
12. Zhou, B.; Lapedriza, A.; Khosla, A.; Oliva, A.; Torralba, A. Places: A 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 1452–1464. [\[CrossRef\]](#)
13. Tang, P.; Wang, H.; Kwong, S. G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition. *Neurocomputing* **2017**, *225*, 188–197. [\[CrossRef\]](#)
14. Li, H.; Cui, Z.; Zhu, Z.; Chen, L.; Zhu, J.; Huang, H.; Tao, C. RS-MetaNet: Deep Metametric Learning for Few-Shot Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6983–6994. [\[CrossRef\]](#)
15. Peng, J.; Tang, B.; Jiang, H.; Li, Z.; Lei, Y.; Lin, T.; Li, H. Overcoming Long-Term Catastrophic Forgetting through Adversarial Neural Pruning and Synaptic Consolidation. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–14. [\[CrossRef\]](#)
16. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1194–1206. [\[CrossRef\]](#)
17. Obeso, A.M.; Vázquez, M.S.G.; Acosta, A.A.R.; Benois-Pineau, J. Connoisseur: Classification of styles of Mexican architectural heritage with deep learning and visual attention prediction. In Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, Florence, Italy, 19–21 June 2017; ACM: New York, NY, USA, 2017.
18. Lee, S.; Maisonneuve, N.; Crandall, D.; Efros, A.A.; Sivic, J. Linking past to present: Discovering style in two centuries of architecture. In Proceedings of the 2015 IEEE International Conference on Computational Photography (ICCP), Houston, TX, USA, 24–26 April 2015; IEEE Computer Society: Los Alamitos, CA, USA, 2015; pp. 1–10. [\[CrossRef\]](#)
19. Zeppelzauer, M.; Despotovic, M.; Sakeena, M.; Koch, D.; Döller, M. Automatic prediction of building age from photographs. In Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, Yokohama, Japan, 11–14 June 2018; pp. 126–134.
20. Zhang, F.; Wu, L.; Zhu, D.; Liu, Y. Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. *ISPRS J. Photogramm. Remote Sens.* **2019**, *153*, 48–58. [\[CrossRef\]](#)
21. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. Deep learning the city: Quantifying urban perception at a global scale. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Cham, Switzerland, 2016; pp. 196–212.
22. Guan, W.; Chen, Z.; Feng, F.; Liu, W.; Nie, L. Urban perception: Sensing cities via a deep interactive multi-task learning framework. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2021**, *17*, 1–20. [\[CrossRef\]](#)
23. Wang, R.; Liu, Y.; Lu, Y.; Zhang, J.; Liu, P.; Yao, Y.; Grekousis, G. Perceptions of built environment and health outcomes for older Chinese in Beijing: A big data approach with street view images and deep learning technique. *Comput. Environ. Urban Syst.* **2019**, *78*, 101386. [\[CrossRef\]](#)
24. Quercia, D.; O'Hare, N.K.; Cramer, H. Aesthetic capital: what makes London look beautiful, quiet, and happy? In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing—CSCW'14, Baltimore, MD, USA, 15–19 February 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 945–955. [\[CrossRef\]](#)
25. Zhou, B.; Liu, L.; Oliva, A.; Torralba, A. Recognizing city identity via attribute analysis of geo-tagged images. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 519–534.
26. Bayram, B.; Kilic, B.; Özoğlu, F.; Erdem, F.; Bakirman, T.; Sivri, S.; Bayrak, O.C.; Delen, A. A Deep learning integrated mobile application for historic landmark recognition: A case study of Istanbul. *Mersin Photogramm. J.* **2020**, *2*, 38–50.
27. Zhang, F.; Zhang, D.; Liu, Y.; Lin, H. Representing place locales using scene elements. *Comput. Environ. Urban Syst.* **2018**, *71*, 153–164. [\[CrossRef\]](#)
28. Zhang, F.; Zhou, B.; Ratti, C.; Liu, Y. Discovering place-informative scenes and objects using social media photos. *R. Soc. Open Sci.* **2019**, *6*, 181375. [\[CrossRef\]](#)
29. Redi, M.; Crockett, D.; Manovich, L.; Osindero, S. What Makes Photo Cultures Different? In Proceedings of the 24th ACM International Conference on Multimedia—MM'16, Amsterdam, The Netherlands, 15–19 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 287–291. [\[CrossRef\]](#)
30. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576.
31. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1510–1519. [\[CrossRef\]](#)

32. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 818–833.
33. Ning, D. PMMS: A Photo based Metadata Mining System for Tourism Research. *Tour. Hosp. Prospect.* **2017**, *1*, 34–47.
34. Hu, Y.; Gao, S.; Janowicz, K.; Yu, B.; Li, W.; Prasad, S. Extracting and understanding urban areas of interest using geotagged photos. *Comput. Environ. Urban Syst.* **2015**, *54*, 240–254. [\[CrossRef\]](#)
35. Yuehao, C.; Long, Y.; Yang, P. City Image Study Based on Online Pictures: 24 Cities Case. *Planners* **2017**, *33*, 61–67.
36. Kita, K.; Kidziński, Ł. Google street view image of a house predicts car accident risk of its resident. *arXiv* **2019**, arXiv:1904.05270.
37. Saleesses, P.; Schechtner, K.; Hidalgo, C.A. The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE* **2013**, *8*, e68400.
38. Hollandi, R.; Szkalitsky, A.; Toth, T.; Tasnadi, E.; Molnar, C.; Mathe, B.; Grexa, I.; Molnar, J.; Balind, A.; Gorbe, M.; et al. nucleAIzer: A parameter-free deep learning framework for nucleus segmentation using image style transfer. *Cell Syst.* **2020**, *10*, 453–458. [\[CrossRef\]](#)
39. Matzen, K.; Bala, K.; Snavely, N. Streetstyle: Exploring world-wide clothing styles from millions of photos. *arXiv* **2017**, arXiv:1706.01869.
40. Shen, X.; Efros, A.A.; Aubry, M. Discovering visual patterns in art collections with spatially-consistent feature learning. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9270–9279. [\[CrossRef\]](#)
41. Karayev, S.; Trentacoste, M.; Han, H.; Agarwala, A.; Darrell, T.; Hertzmann, A.; Winnemoeller, H. Recognizing image style. *arXiv* **2013**, arXiv:1311.3715.
42. Thomee, B.; Shamma, D.A.; Friedland, G.; Elizalde, B.; Ni, K.; Poland, D.; Borth, D.; Li, L.J. YFCC100M: The new data in multimedia research. *Commun. ACM* **2016**, *59*, 64–73. [\[CrossRef\]](#)
43. Kalkowski, S.; Schulze, C.; Dengel, A.; Borth, D. Real-time analysis and visualization of the YFCC100M dataset. In Proceedings of the 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions—MMCommons’15, Brisbane, Australia, 26–30 October 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 25–30. [\[CrossRef\]](#)
44. Howard, J.; Ruder, S. Universal language model fine-tuning for text classification. *arXiv* **2018**, arXiv:1801.06146.
45. Do, C.B.; Ng, A.Y. Transfer learning for text classification. *Adv. Neural Inf. Process. Syst.* **2005**, *18*, 299–306.
46. Zhu, Y.; Chen, Y.; Lu, Z.; Pan, S.J.; Xue, G.R.; Yu, Y.; Yang, Q. Heterogeneous transfer learning for image classification. In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, CA, USA, 7–11 August 2011.
47. Quattoni, A.; Collins, M.; Darrell, T. Transfer learning for image classification with sparse prototype representations. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; IEEE: Los Alamitos, CA, USA, 2008; pp. 1–8. [\[CrossRef\]](#)
48. Lim, J.J.; Salakhutdinov, R.; Torralba, A. Transfer Learning by Borrowing Examples for Multiclass Object Detection. In *Advances in Neural Information Processing Systems 24: Proceedings of the 25th Annual Conference on Neural Information Processing Systems 2011, Granada, Spain, 12–14 December 2011*; Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F.C.N., Weinberger, K.Q., Eds.; Massachusetts Institute of Technology: Cambridge, MA, USA, 2011; pp. 118–126.
49. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Huh, M.; Agrawal, P.; Efros, A.A. What makes ImageNet good for transfer learning? *arXiv* **2016**, arXiv:1608.08614.
51. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988. [\[CrossRef\]](#)
52. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2921–2929. [\[CrossRef\]](#)
53. Shamsuddin, S.; Sulaiman, A.B.; Amat, R.C. Urban landscape factors that influenced the character of George Town, Penang UNESCO World Heritage Site. *Procedia-Soc. Behav. Sci.* **2012**, *50*, 238–253. [\[CrossRef\]](#)
54. Cheng, L.; Chu, S.; Zong, W.; Li, S.; Wu, J.; Li, M. Use of tencent street view imagery for visual perception of streets. *ISPRS Int. J.-Geo-Inf.* **2017**, *6*, 265. [\[CrossRef\]](#)