Using Flickr Data to Understand Image of Urban Public Spaces with a Deep Learning Model: A Case Study of the Haihe River in Tianjin

Chenghao Yang, Tongtong Liu, Shengtian Zhang

Department of Architecture, School of Architecture, Tianjin University, Tianjin 300072, China
Department of Architecture, College of Design and Engineering, National University of Singapore, Singapore 117566, Singapore
School of Computer and Network Engineering, Shanxi Datong University, Datong 037009, China

Abstract: Understanding public perceptions of images of urban public spaces can guide efforts to improve urban vitality and spatial diversity. The rise of social media data and breakthroughs in deep learning frameworks for computer vision provide new opportunities for studying public perceptions in public spaces. While social media research methods already exist for extracting geo-information on public preferences and emotion analysis findings from geodata, this paper aims at deep learning analysis by building a VGG-16 image classification method that enhanced the research content of images without geo-information. In this study, 1940 Flickr images of the Haihe River in Tianjin were identified in multiple scenes with deep learning. The regularized VGG-16 architecture showed high accuracies of 81.75% for the TOP-1 and 96.75% for the TOP-5 and Grad-CAM visualization modules for the interpretation of classification results. The result of the present work indicate that images of the Haihe River are dominated by skyscrapers, bridges, promenades, and urban canals. After using kernel density to visualize the spatial distribution of Flickr images with geodata, it was found that there are three vitality areas in Haihe River. However, the kernel density result also shows that judging spatial visualization based solely on geodata is incomplete. The spatial distribution can be used as an assistant function in the case of the under-representation of geodata. Collectively, the field of how to apply computer vision to urban design research was explored and extended in this trial study.

Keywords: deep learning; image classification; urban public space; social media data; Flickr

1. Introduction

Urban public space is directly related to the quality of life of inhabitants and the attractiveness of tourism, and images of urban public spaces are highly consistent with images of the central areas of cities. Public space is not only a place for providing public activities to citizens and tourists but also a carrier of urban culture [1]. Many popular public spaces have become synonymous with their cities [2], such as Central Park, Bryant Park, and Times Square in New York and the Riverfront in Chicago. There is a consensus in the study of public spaces that diverse urban public life is the core concept of creating public space and that visual perception dominates the subjective feeling of urban public life. Sociological topics such as the privatization of public space [3], gentrification, social control [4], public participation, and social justice in public spaces are the most popular research topics. Jane Jacobs thinks that only democratically shared public spaces can truly generate public life [5]. No matter the direction of the related research on public space, people-oriented spaces and public participation have always been the most important topics [6–8].

Public space surveys aim to improve the quality of public spaces and improve the public lives of residents and tourists. Aiming at the relationship between urban life and
public space, Jan Gehl gradually formed a complete public space and public life (PSPL) survey after years of research [9]. However, the traditional research on images of urban public spaces is hindered by limited investigation funds and time, investigators, and other conditions and cannot be sustained over long periods.

In recent years, the proposition of using social media data to investigate and optimize urban space has received extensive attention from scholars, and deep learning methods have broad prospects in the field. Although deep learning techniques are increasingly used in urban and landscape research, research on urban design scales such as urban public space remains to be developed. With the development of artificial intelligence and deep learning, urban public space research is not limited to conventional environmental psychology approaches. Based on the progress of computer vision (CV) and location-based social media (LBSM) technology and the popularization of corresponding service products, the views of urban public space by locals and tourists in various periods can be learned in real time if there are sufficient data [10]. The most practical advantage of the new technology lies in mapping urban images and urban spatial vitality through volunteered geographic information (VGI) data [11]; the geodata can reflect the behavioral characteristics and spatial distributions of public preferences [12]. Since public space placemaking is an ongoing task, minor tweaks can be made to improve the usefulness of urban public spaces over time, and deep learning analysis feeds into the dynamic process of public life change.

However, the previous studies using deep learning for image classification have mainly analyzed visual perception and have rarely combined the user’s location, the geodata distributions of image classifications, and current land uses to provide placemaking suggestions for urban public spaces. To fill this gap and discover a valuable method for understanding images of urban public spaces, this study proposed using deep learning image classification for the deep excavation of images of urban public spaces and to provide several placemaking suggestions for retrofitting existing urban public spaces or planning new spaces based on kernel density and qualitative analysis results. Furthermore, this study aimed to analyze which place scenes receive more attention through Flickr images, what kinds of people are posting the images, where these preferred scenes are spatially distributed, and what recommendations the results suggest for placemaking.

The remainder of this paper is organized as follows. Section 2 discusses related work on Flickr data in the field of urban research and the CV method based on deep learning. In Section 3, we introduce the study area with its historical development and the current situation of the Haihe River urban public space in Tianjin; we also introduce descriptive metadata and image datasets from the Flickr API interface and describe how we obtained and cleaned them. Meanwhile, we present our analytic approach metadata and mapping visualization and image analysis methods using Visual Geometry Group Net with depth of 16 layers (VGG-16) based on convolutional neural network (CNN) deep learning in TensorFlow 2.0 and Keras framework. In Section 4, results are presented from three main study tasks. The first was the analysis of the Flickr metadata. The second was to classify the Flickr images using optimized VGG-16 deep learning. The final component was to use the kernel density for intensity and distribution of vitality, combined with current land use for qualitative analysis. In Section 5, we discuss the advantages and disadvantages of using social media data to measure images of urban public spaces and put forward three urban design placemaking suggestions for Haihe’s urban public spaces as well as urban generation guidance for future work. We conclude, in Section 6, with a summary of our main method for investigating and analyzing images of urban public spaces.

2. Related Work

Recent years have witnessed the fast development and popularity of using big data for the comprehensive measurement of human behavior including public perceptions of urban spaces [13]. Wi-Fi probe and location data [14], GPS device tracking [15], mobile phone data [16], and social media data [17] are gradually being applied in research on public spaces. It can be said that big data is being harnessed to satisfy changing public needs.
Meanwhile, dynamic data are replacing static data, and multi-source data will replace single-source data. For this study, using traditional urban design methods to measure images of urban public spaces can only choose a few days for observing the changes in urban images, but deep learning based on big data research can explain more information and costs less. In general, while the use of social media data in urban studies research is still in its infancy, the field coverage is gradually increasing in multiple directions, especially in the fields of CV and natural language processing.

In urban studies using CV and LBSM data, the main social media platforms studied are Twitter, Facebook, Flickr, Foursquare, and Sina Weibo, which are the most popular social media platforms based on web 2.0. Among them, Twitter and Flickr are the most frequently studied platforms [18,19]. In social media research in urban studies, Flickr data are the second most frequently used data after Twitter [20]. Flickr data research focuses on metadata and image data. Flickr metadata can reflect spatial visitation patterns and their features [21], visitor provenance and patterns of recreation [22], visit frequency [23], spatial distributions [24], landscape aesthetics [25], and urban areas of interest [26]. Metadata research is highly dependent on the number and accuracy of geo-information and geotagged images, which are based on 100,000-level data volumes. Social media photos include those with human beings and scenes, the former reflecting human activities and the latter recording users’ appreciation of spaces or elements [27].

In the CV research using Flickr data, clustering, traditional machine learning, and deep learning are the mainstream methods for extracting images, classifying images, and analyzing potential information. Gosal [28] and Wartmann [29] et al. used Google’s Cloud Vision APIs and Python scripting for photo retrieval and content analysis, which proved that Cloud Vision is an effective image analysis tool. Ashkezari-Toussi [30] analyzed the city’s emotional structure through the expressions of human beings in social media images. In addition, several studies have been conducted to analyze urban image or landscape aesthetics by applying deep learning pretrained models based on ImageNet and Places365 datasets. Kim aimed to show representative urban imagery by analyzing Seoul’s Flickr photos using the Inception v3 model in the ImageNet dataset. Kang [31] continued with the Inception v3 model to analyze the regions of attraction and the tourists’ urban image in Seoul using transfer learning and improved the top-1 accuracy to 85.77% and the top-5 accuracy to 95.69%. Some studies clearly pointed out the problem of low classification success rates [32]. Therefore, finding suitable image classification architecture for the Haihe River can improve the accuracy rate of Flickr data, such as ResNet, GoogLeNet, and VGG-16 based on CNN.

3. Materials and Methods

3.1. Study Area

The Haihe River has a total length of 73 km; it is the largest river in North China. The Haihe River is not only a symbol of urban culture in Tianjin but also the most attractive urban space for residents and tourists. The core area of Tianjin’s comprehensive Haihe River public space is located near Tianjin Station, which is the historical and cultural resource of lamps and the business center in the central urban area of Tianjin (Figure 1). In the core area of Haihe River, on the north side of the river are Tianjin Italian Style Town, office buildings, hotels, and Tianjin Station, and on the south side of the river are hotels and apartments, a library, office buildings, a theater, and the commercial center: Jinwan Plaza. In this area, there are riverwalks near Haihe, a waterfront square, multiple parks, and landscape sketches to form a water-friendly landscape belt.
Plaza. In this area, there are riverwalks near Haihe, a waterfront square, multiple parks, and landscape sketches to form a water-friendly landscape belt.

Figure 1. Core area and nodes of Haihe River.

Tianjin began the urban design of both sides of the Haihe River in the 1990s and formally started the comprehensive redevelopment of the river in 2002 with public participation [33]. In addition to the renewal of Haihe River on both sides of riversides, the urban design includes the treatment of industrial workshops and river sewage along the river, the protection and reuse of buildings with high cultural value along the river, and the urban renewal of residential areas and service industry along the river [34]. At the same time, the Haihe River not only affects the spatial vitality along the river but also catalyzes the vitality of the whole city.

Just like the Seine River in Paris [35], the Thames River in London, and the Yarra River in Melbourne, the Haihe River carries the daily lives of residents and is the center of urban culture and public life. Citizens need the public space to sustain more efficient production vitality, their social cohesion and inclusion, their civic self-identification, and their quality of life. Our research focuses on understanding images of Haihe River, which can reflect the state of neighborhood community life and tourism preference.

3.2. Data Collection

Flickr, established in 2004 in Canada, is an image and video hosting social media platform with over 112 million registered members worldwide. Compared with balanced text comments and sharing images on social media platforms, Flickr is a social media platform mainly based on image sharing. Some Flickr images come with geographic coordinates, which provide for spatial analysis based on image metadata. Geolocated Flickr photographs have been used effectively to quantify nature-based tourism and recreation worldwide [36]. In urban studies research, Twitter and Flickr are both dominant data sources, while an increasing number of studies from China focus on Sina Weibo [20]. Some studies reveal the fact that Flickr is mainly used in North America and Europe, and it is not particularly popular among Chinese users [37,38]. Therefore, in terms of applicability,
Weibo seems to be a more suitable social media platform than Flickr for studying Chinese cities. As Weibo’s location information service is currently access by commercial API, Flickr is still more appropriate for this study.

Flickr’s data are obtained from the Flickr API (https://www.flickr.com/services/api/ (9 February 2022)) after applying for the API OAuth key. The API function used in this research is the “flickr.photos.search” function, which can be accessed by sending a REST request to the endpoint. There are three ways to search for Flickr images: The first is based on Places ID, the second is a radius and bbox (rectangle) search based on a certain coordinate, and the third is a keyword search. Keyword search is used in this study, which can maximize the capture of shared images without geo-tags. The keyword is English “Haihe” and Chinese “Haihe”. The data are divided into two categories: “has geo” and “without geo”. After cleaning up data such as “Hai River” or “Haihe” in other cities in local time (GMT+8, Beijing Singapore) and removing duplicate images and images not related to the Haihe River, there were 1940 Flickr images, including 440 images with metadata with geographic information, and 1500 images without metadata. This returned response in specific CSV format metadata and JPG format image data. To comply with Flickr API terms and privacy policies, all metadata and photo owner data sets were anonymized, and unnecessary personal information data were cleaned.

Since only 18.9% of the Flickr images have high-precision geodata, this study used the perspective of Flickr image feeds as the source of the geo-tagged data. The visualization results also represent the spatial distribution of public attention along the Haihe River. Table 1 shows examples of Flickr data with geo-information and non-geodata metadata and images. To protect the privacy of user images, the last five digits of the image ID are hidden. The copyright of each Flickr image belongs to the photographer.

### Table 1. Example of Haihe River Flickr image data obtained through the Flickr API.

| Image ID   | User’s Location | Take Date       | Geo-Location | Sharing Images |
|------------|-----------------|-----------------|--------------|----------------|
| 490632 ***** | China           | 21 August 2018  | 39.13093, 117.198819 |                |
| 119855 ***** | USA             | 17 December 2013| 39.40995, 117.38047 |                |
| 47713 ***** | USA             | 23 June 2010    | Non-geodata  |                |
| 15072 ***** | Tianjin         | 7 October 2007  | Non-geodata  |                |

3.3. Data Analysis

The flowchart of the method is shown in Figure 2. Based on the Flickr data metadata and images, this study required three phases to analyze the Flickr data. Phase 1 was the Flickr grouping analysis, including tourists and locals, age and gender, and machine tags. Phase 2 was the Flickr image classification, including establishing VGG-16 architecture,
using the L2 regularization and dropout regularization tools after using the VGG-16 architecture and Grad-CAM (Gradient-weighted class activation mapping) visualization after the output of the classification result. Phase 3 was the visual analysis of the spatial strength and spatial distribution of the urban public spaces. This stage consisted of the kernel density analysis of two-dimensional public spaces based on ArcGIS Pro.

Figure 2. Flowchart of the research.

The results of Phase 1 reflect the changes in the posting dates, and user’s locations for when the Flickr images were shared. It was possible to analyze whether the user was a tourist or a resident based on their location. Phase 2 was the core of the research, and the image classification could reflect the overall image of the Haihe River. It could also be interpreted as reflecting the public’s scene preferences. Phase 3 was a further interpretation of the urban public space imagery that could obtain the spatial distributions of public spatial intention.

Based on the kernel density results, there were several qualitative analyses for urban public space suggestions, including the discovery of hotspot areas on the Haihe River, the comparison of the geodata image classification results with current land development, and the comparison of the image classification results between geodata and non-geodata images. The research constitutes a methodology for understanding the preferred images of urban public spaces based on social media image data and metadata. The analysis results also lead to three urban public space placemaking suggestions for the Haihe River.

3.3.1. Grouping Statistics

In addition to geotags, metadata such as the name, hometown, and sharing time of each Flickr image sharer can be obtained from the Flickr API. Here, we counted the taken date of the Flickr image and the user location. We grouped the time slices of image data. There were 1940 Flickr images shared of the Haihe River, which is a small data set. Using the Places database to train an image classification model suitable for
public space scene recognition could solve the problem of low accuracy in small data. The Places365 dataset was the most suitable dataset for this research, designed by the MIT Computer Science and Artificial Intelligence Laboratory. The dataset contains 5000 training images per class in a total of 365 classes [39]. ImageNet is an object-centric database from which it is hard to extract public space scene labels, making it necessary to also train on Places365 [32]. Using the deep learning CNN framework, these image data were divided into multiple classes for this research objective according to the Places365 basic classes.

Several studies have proved that VGG-16 is reliable in top-1 accuracy for image classification tasks [40–42], and it shows the highest top-1 accuracy for the Places365 database. The prominent benefit of the VGG16 model is that it enhances the performance of CNNs without the necessity of deeper training with a high number of convolutional layers [43]. This study built the VGG-16 model in the Keras TensorFlow framework. The core structures of CNN such as convolutional layers, pooling layers, and fully connected layers are provided by Keras corresponding functions. The process can be divided into six main steps (Figure 3).

![Figure 3. VGG-16 architecture for Image classification framework.](image)

1. Download the Places365 basic dataset, including training set (5000 per class, 1,825,000 in total), validation set (100 per class, 36,500 in total), and test set (900 per class, 328,500 in total).
2. Set the image size of the Places365 standard dataset to 224 × 224 × 3 and use it to train, evaluate, and test on the optimized VGG-16 model to obtain a classifier.
3. Use the Keras TensorFlow framework to construct the corresponding structural components of the VGG16 model, including convolutional layers, maximum pooling layers, and fully connected layers.
4. To prevent overfitting, L2 and use dropout regularization based on the original VGG-16 model.
5. Use the optimized VGG-16 model as a classifier to classify the Flickr images, with each image corresponding to 365 predicted label probabilities. Sort the predicted label probabilities in descending order and select the top five probabilities as the top-1 to top-5 image output results.
6. Use Grad-CAM to visually interpret Flickr images and identify and locate relevant image regions. Grad-CAM uses the gradient of network back-propagation to calculate...
the weight of each channel of the feature map to obtain the corresponding heat map [44].

Grad-CAM can be applied to a variety of different image tasks.

The architecture model can realize the function of the scene recognition classifier with high precision and high recall rate, can provide corresponding scene feature labels for each Flickr image, and can calculate the predicted categories and weights of top-1 to top-5. Among the output results, the prediction results for top-1 to top-5 accuracy are particularly important. Top-N accuracy is a measure of how often the predicted class falls within the top N values of the SOFTMAX distribution. In ImageNet, the error rate is often used to explain the probability of image recognition [45,46]. Top-1 accuracy is the conventional accuracy: the model response must be exactly the expected answer. Top-5 accuracy means that any of the five highest probability results must match the expected response.

3.3.3. VGG-16 Architecture and Grad-CAM

Compared with the deep learning model, the traditional machine learning image classification model has lower recognition; CNNs can achieve better image classification accuracy on large-scale datasets [47,48]. Deep learning models can generally be improved to more than 70% top-1 accuracy; for instance, Amazon’s team raised ResNet-50’s top-1 validation accuracy from 75.3% to 79.29% on ImageNet in 2019 [49], which will be beneficial for obtaining more image classification information from image data. In addition, support vector machines are also effective classifiers for scene classification tasks [50].

In our study, VGGnet is an optimization model based on the CNN network, proposed by Karen Simonyan and Andrew Zisserman from the University of Oxford in 2014. In our research, we used the VGG-16 architecture for image classifiers [48]. VGG-16 has 16 weight layers (‘16’ or ‘19’ stand for the number of weight layers in the network), including 13 convolutional layers, 3 fully connected layers, and 5 pooling layers. Both variants, VGG-16 and VGG-19, consist of two fully connected layers; all the conv kernels are 3 × 3; and the maxpool kernels are 2 × 2 (Smaller than AlexNet’s 3 × 3 pooling kernel) with a stride of two, each with 4096 channels, followed by another fully connected layer. The convolution kernel focuses on expanding the number of channels, and the pooling focuses on reducing the width and height; the model architecture is deeper and wider. Unfortunately, VGGNet has two major drawbacks which are slow training speed and large architecture weights (16 or 19 weights). VGG-16 was built in Keras TensorFlow layers. Although there were more efficient networks like ResNet and SENet architecture in ILSVRC for image classification, VGG-16 converges and is then used as initialization for larger, deeper networks, a process called pretraining. In summary, we believe that VGG was a more suitable architecture based on the amount of data in urban public space research for small and medium urban scales.

Overfitting is a common problem of neural network approaches. To alleviate the overfitting problem of VGG-16 in the training process, this study applied L2 regularization and dropout regularization based on the original VGG-16 architecture in TensorFlow Keras. L2 regularization is also called weight decay. The principle is to increase the L2 regularization term based on the original loss function, thereby restraining the original loss function and suppressing overfitting caused by excessive weight [51].

Using the L2 regularization method of Keras Regularizers object, we added L2 regularization to the convolutional layer with the help of the Kernel_regularizer keyword argument of the convolutional layer, with a regularization factor of 0.0002.

The L2 regularization function applied to the original loss function is:

$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} L\left(g^{(i)}, y^{(i)}\right)$$  \hspace{1cm} (1)
In the formula: \( m \) represents the numbers of input batch data, \( \hat{y}^{(i)} \) is the \( i \)th predicted value, \( y^{(i)} \) is the \( i \)th target value, and \( L(\hat{y}^{(i)}, y^{(i)}) \) is the loss value of the \( i \)th data. Suppose the derivative of the original loss function concerning \( w \) is:

\[
\frac{\partial J(w^{[l]})}{\partial w^{[l]}} = dw^{[l]}
\]  

(2)

\( l \) represents the neural network layer. Then the weight update formula is:

\[
w^{[l]} = w^{[l]} - \alpha dw^{[l]}
\]  

(3)

After adding the regularization term, the loss function is changed to:

\[
J(w, b) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) + \frac{\lambda}{2m}\|w^{[l]}\|_2^2
\]  

(4)

In the formula: \( \lambda \) is the regularization coefficient. The new loss function takes the derivative of \( w \):

\[
\frac{\partial J(w^{[l]})}{\partial w^{[l]}} = dw^{[l]} + \frac{\lambda}{m}w^{[l]}
\]  

(5)

This means the new weight update formula is obtained as:

\[
w^{[l]} = w^{[l]} - \alpha \left( dw^{[l]} + \frac{\lambda}{m}w^{[l]} \right)
\]  

(6)

\[
w^{[l]} = w^{[l]} \left( 1 - \alpha \frac{\lambda}{m} \right) - \alpha dw^{[l]}
\]  

(7)

It can be seen from Formula (7) that \( 1 - \alpha \frac{\lambda}{m} \) is a number that is less than 1, which will reduce \( w \), reduce the influence of \( w \) on the neural network, and alleviate the problem of overfitting.

Dropout is a technique for improving neural networks by reducing overfitting [52]. The dropout for regularizing is to randomly delete neurons in the neural network so that the network has a certain degree of sparseness, effectively reducing the synergistic effect of different features. It weakens the joint adaptability between neuron nodes, increases the generalization ability, and effectively reduces the occurrence of overfitting. The forward propagation calculation formula of the neural network without dropout is:

\[
z^{(l+1)}_i = w^{(l+1)}_i y^{(l)} + b^{(l+1)}_i
\]  

(8)

\[
y^{(l+1)}_i = f(z^{(l+1)}_i)
\]  

(9)

The neural network forward propagation calculation formula after adding the dropout regularization is:

\[
r_i^{(l)} \sim Bernoulli(p)
\]  

(10)

\[
\hat{y}^{(l)} = r^{(l)} \times y^{(l)}
\]  

(11)

\[
z_i^{(l+1)} = w^{(l+1)}_i \hat{y}^{(l)} + b^{(l+1)}_i
\]  

(12)

\[
y_i^{(l+1)} = f(z^{(l+1)}_i)
\]  

(13)

The Bernoulli function in Formula (10) was used to randomly generate a vector of 0 and 1 with probability \( p \). In this study, the dropout was introduced into the two-layer fully connected network in the VGG16 model, and the dropout probability was set to 0.5 because dropout randomly generates the most networks when the dropout probability is
0.5 [53]. At this point, we had obtained an optimized VGG-16 architecture for predicting image classification. The optimized image classifier included VGG-16, L2 regularization acting after 13 convolutional layers, and dropout regularization for 2 dense layers.

While deep learning has facilitated unprecedented accuracy in image classification, one of the biggest problems is model interpretability, a core component in model understanding and model debugging. To make the results of the image classification interpretable, the Grad-CAM [54] visualization method was used in this study to obtain heatmaps that characterized image classification. Grad-CAM works by finding the convolutional layer in the VGG-16 network and examining the gradient information flowing into that layer after output results. Grad-CAM is an upgraded version of CAM [54] (class activation mapping), which does not need to modify the neural network and retrain it.

3.3.4. Mapping the Image of Urban Public Space with Geodata

Among visualization methods, kernel density is one of the most popular and well-established techniques, and it is the most used in mainstream urban spatial research because ArcGIS pro. kernel density analysis (Spatial Analyst) provides quantitative values and the ability to visualize the concentration of points or lines. According to Silverman [55], kernel density calculates the density of point features around each output raster cell [56]. The kernel density estimated the utilization distribution in each pixel of a grid superimposed on the locations of an individual’s sharing data [57].

In related studies, kernel density analysis has been widely used to detect patterns of continuity and discontinuity in social media activities [58]. The mapping result here showed the spatial agglomeration and distribution characteristics of the Flickr metadata with geoinformation. In our research, the kernel density analysis generated continuous surfaces using Flickr metadata to visualize where the point data sets were. Most kernel density research used urban and regional scales with radii of 500 m, 1000 m, or 5000 m. As for the urban design scale, this method was influenced by the seminal works on pedestrianized areas [59] and public open spaces [60]. Smaller preset values produce a raster (output cell size) that shows more detail of spatial vitality, which means that only high-precision kernel density mapping is suitable for public space scale research [61] which is performed using the formula below:

$$f(x; h) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)$$

To measure the density of the Flickr metadata points at point $x$, where $X_i$ is the Flickr dataset containing $X$ (latitude) and $Y$ (longitude) coordinates and $X_i = (X_i, Y_i)^T$, and $i = 1, 2, 3, \ldots, n$. The $h$ is the bandwidth, also known as a smoothing parameter. As long as the bandwidth is determined, the influence of kernel functions in various mathematical forms on kernel density is small [62].

4. Results

4.1. Metadata Statistics

According to the statistics from the Flickr API, metadata from a total of 1940 images was obtained. After cleaning the duplicate Flickr images, the module of taken_date was particularly noteworthy, reflecting people’s preferences in different periods. Unfortunately, due to the small scale of the public space, the amounts of metadata were small, and the time slices could can only be set to the years 2017, 2013, and 2009; these were the peaks of the Haihe Flickr metadata (Figure 4). These peaks might have been because the holding of the 2008 Olympic Games and the Davos World Economic Forum in Tianjin (2010, 2012, 2018) attracted more tourists to the Haihe River. More importantly, this is inseparable from the urban design project on the Haihe River, which has been under continuous construction since 2008. In addition, because the COVID-19 pandemic reduced the sharing of Flickr images, the existing data also gradually decreased beginning in 2020. From the comparison of geodata and non-geodata images, geographic data images are much rarer than images without geographic coordinates.
The metadata also includes the user’s information, including 1393 data with the user’s location and another 547 with unknown user location data. Statistically (Figure 5), users in North America shared the most Flickr images, with data with geographic coordinates accounting for 22% and coordinates without geo-information accounting for 46% of the total data. The data shared by local users accounted for 4% with geo-information and 21% without geo-information of the total data. Flickr data shared by domestic visitors also stood out, with geo-information representing 33% of the total data and without geo-information representing 12% of the total data. Furthermore, users from Asia and Europe also maintained a high level of interest in photographing the Haihe River. Collectively, a high percentage of Flickr images were from domestic and overseas tourists, with most Flickr images shared by residents not containing geodata. In contrast, the spatial distribution of geodata better reflected the tourists’ image preferences for urban public spaces, especially overseas tourists, and the low proportion of geographic data from residents affected the distribution judgment for local groups. Although the proportion of local residents is relatively small, this study can still reflect residents’ and tourists’ perceptions.

4.2. Image Classification Result
Before training the VGG-16 deep learning architecture, we predicted the image scene type in 1940 images utilizing manual classification. We believe that there should be at
least eight types of scenes in Places365, with the highest proportions centered on riverfront public spaces like the Haihe River, namely: bridge, canal/urban, downtown, ice floe, park/water park, plaza, promenade, and skyscraper (Figure 6).

Table 2 shows an example of the inconsistencies between the manual labels and prediction results. The misrecognized images were mostly underexposed or overexposed. This optimized VGG-16 achieved top-5 test accuracy of 96.75%. Due to some ambiguity between public space scene categories, and the features of multiple scenes appearing in the same image, top-5 accuracy is the standard to measure the performance of the image classification model [39,49].

Table 2. Examples of Top-1 and Top-5 Error classification.

| VGG-16 Results | Results Comparison |
|----------------|--------------------|
| Top-1 Error     | Manual Classification: Skyscraper |
| Top-5 Correct   | VGG-16 results: Industrial area, skyscraper, downtown, office building, tower |
| Top-1 Error     | Manual Classification: Bridge |
| Top-5 Error     | VGG-16 results: Skyscraper, downtown, tower, industrial area, office building |

To verify the top-1 and top-5 accuracy, two experienced researchers randomly selected 400 Flickr images of the Haihe River for manual classification. The results show that the manual classification labels matching the top-1 prediction results accounted for 81.75% of the total 400 images, and the Flickr images matching the top-5 prediction results accounted for 96.75% of the total 400 images.

Figure 7 shows the example output result as true. For example, image A is canal/urban, and the predicted top-1 to top-5 results were canal/urban, skyscraper, downtown, industrial.
area, and harbor. Image E is Bridge, and predicted top-1 to top-5 results were bridge, skyscraper, downtown, tower, and river. The prediction result also shows the probability of the image classification label. Since the dataset Places365 has 365 categories, there are 365 probability values predicted for each image. From the perspective of prediction probability, the top-1 prediction probability of the image A to image F is 12.99%, 16.23%, 23.74%, 35.78%, 33.6%, 64.4%, respectively. The predicted result is highly consistent with the real label category: canal/urban. In contrast, the predicted result may not always match the real label, even within the possibility of top-1–top-5 performance; we defined such cases, like image B, as errors. Unlike other urban-scale studies, the image of urban public spaces in Haihe have a high degree of similarity, which increased the difficulty of this image recognition; the top-5 accuracy became a perfect aid.

**Figure 7.** Prediction results of VGG-16 architecture for Flickr images.

The principle of Grad-CAM is to obtain partial derivatives of all feature maps of the last convolutional layer according to the node probability with the largest SOFTMAX value and then select the mean of the gradient of each feature map as the weight to obtain a weight vector [63]. Multiply the weight vector and the feature map correspondingly and add them to obtain a two-dimensional matrix. Then the two-dimensional matrix is sent to ReLU for activation [64], and the negative numbers in the two-dimensional matrix are changed to 0. Finally, sampling is performed to obtain the Grad-CAM heatmap and to combine Grad-CAM with guided backpropagation (Figure 8).
In general, after counting the top-1 prediction results (prediction correct), the most images were in the skyscraper (321), bridge (299), promenade (155), canal/urban (120), downtown (107), harbor (66), tower (64), mosque/outdoor (61), and ice floe (35) categories (Figure 9). The prediction results are consistent with the conjecture results before the deep learning classification experiment. The skyscrapers in the central core area of the Haihe River and the 30 bridges on the Haihe River were the most popular subjects of urban public space imagery. Skyscrapers are mainly concentrated on the south side of Tianjin Station and both sides of Haihe Culture Square, while bridges are evenly distributed at the intersections of main roads with the Haihe River. Among the results, the predictions for the tower and mosque/outdoor scenes were surprising. After we sorted the data, we found that most of the images categorized as mosque/outdoor from the original Flickr images were facades of Tianjin Station and the tower classified was the clock tower of Jinwan Plaza. Figure 10 shows the predicted results from top-1 to top-5, which means that each image has five classification labels: the most numerous image categories are skyscraper (747), downtown (700), bridge (624), tower (531), harbor (456), river (450), promenade (425), office_building (376), and canal/urban (334). The top-5 results are more consistent with the top-1 results. The scenes of downtown, tower, river, etc. are highly similar to the labels of skyscrapers and canal. Overall, the top-1 prediction results are reliable. The image classification led to the following conclusions: the high-rise buildings in the core area of the Haihe River, the bridges, and the promenades along the canal are the most interesting to the public.
To compare the spatial distributions of the image classification results from the VGG-16 architecture, we performed kernel density analysis for the images with geographic coordinates. The areas shaded in red in Figure 11 indicate higher crowd density, higher activity frequency, and higher concentration of social media use [65]. Some features are highly popular from the Flickr geodata mapping in Figure 11, such as the Observation Deck, Beian Bridge, Jiefang Bridge, Tianjin Station Square, Haihe Culture Square, and Jinwan Plaza in the core area of Haihe River. This area provides a wealth of public events for overseas tourists, national tourists, and residents. As an important landmark along the Haihe River, the Tientsin Eye attracts a large number of visitors. It has also led to the construction of surrounding residential, commercial complexes, parks, and the Tianjin Children’s Palace which is under construction. In addition, the kernel density mapping result also shows that many images are concentrated along the Haihe E Road (i.e., red area in the figure). This may be highly related to Haihe Park and the hydrophilic platform here. However, we still consider the concentration of images of the urban public spaces along Haihe E Road to be anomalous because one user took 52 consecutive images on 22 July 2010. This corroborates our expected idea that the results of image classification using only
Only geodata may be unilateral. Combining the results of the previous section, we prefer to rank Haihe E Road third in overall images shared, just below the core area of the Haihe River and the Tientsin Eye.

![Kernel density result of Haihe River](image)

**Figure 11.** Kernel density result of Haihe River.

Based on the kernel density result, the classification results can be divided into three areas where urban public space imagery is highly concentrated. Figure 12 shows the relationships between the image classification results for geodata and current land uses on the Haihe River. Judging from the current land use, the urban functions of Area 1 are mainly commercial, commercial/office, and transportation; of Area 2, mainly residential, commercial/residential, and campus; and of Area 3, mainly commercial/office, commercial/residential, and residential. From the perspective of urban public space functional characteristics, Area 1 is large open spaces combined with commercial centers and transportation, Area 2 prefers community parks and recreational facilities, and Area 3 tends toward community parks, pocket parks, and riverside trails.

The categories with the largest number of top-1 image classification results were skyscraper, bridge, promenade, canal, downtown, and ice floe. Using the Graduated Symbols for Mapping tool, symbol size showed quantitative differences [66]. The skyscrapers and downtown predicted by VGG-16 are concentrated in Area 1 and Area 2. The categories of predicted images that appear in all three areas include bridges, promenade, canal, and harbor. As for the tag ice floe, it is mainly located in Area 2. That indicates that ice activities in winter are mainly concentrated in Area 2.
Figure 12. Spatial analysis of image classification TOP-1 results and current development land uses.

In summary, Area 1 is located within a one-kilometer radius of the Century Clock (near the Jiefang Bridge), the representative images of urban public spaces based on the VGG-16 image classification results are skyscraper, bridge, promenade, downtown, and tower. The result shows that Jinwan Plaza, the office buildings, and their affiliated urban public spaces are the most attractive scenes, with a mixed crowd structure. Area 2 is located within a 300 m radius of the Tientsin Eye, and there are 3 bridges nearby that have attracted a great deal of attention. The Flickr data in this area are mostly from domestic visitors. Water park, harbor, and amusement park are the scenes that people pay attention to. Area 3 is located between the Zhigu Bridge and Dongxing Bridge, and the results are promenade, bridge, canal/urban, harbor, and industrial area. In addition, Area 3 is most densely populated, with residences along both sides of the river in conjunction with several community parks that are mostly used by locals (Table 3).

Table 3. Core area of the Haihe River based on visualization results.

| Core Area of Haihe River | The Highest Image Classification Results | Major Groups      |
|--------------------------|-----------------------------------------|-------------------|
| Area 1                   | Skyscraper, Bridge, Promenade, Downtown, Tower | Overseas tourists  |
|                          |                                         | Domestic tourists |
|                          |                                         | Residents          |
| Area 2                   | Promenade; Bridge, Canal/urban, Harbor, Amusement Park | Domestic tourists  |
|                          |                                         | Overseas tourists  |
| Area 3                   | Promenade, Bridge, Canal/urban, Harbor, Industrial Area | Residents          |

After counting the image classification results for geodata and non-geodata images in Figure 13, the proportion results show that the images of skyscrapers, promenades, and the harbor airport terminal are falsely high. Among them, the label airport terminal should
be ignored, as Tianjin’s airports are not located on either side of the Haihe River. At the same time, the proportion statistics also show that the images of the bridge, canal/urban, downtown, and tower in the geodata are lower than the total proportions. Compared with the classification results for geodata, the proportions of classification results for non-geodata images are closer to the total proportion. Although geodata plays a huge role in spatial analysis, the insights gained from this section may be of assistance to future researchers on non-geodata images.

Figure 13. The proportion of Top 20 image classification results for geodata and non-geodata images.

5. Discussion

As research on sharing images on social media platforms has recently become widely used in urban planning, urban design, and landscape design, using deep learning technology to analyze images has broad prospects for urban studies and physical geography. CV tasks such as image classification, image segmentation, and object detection can reflect public perceptions and visual tendencies. At the same time, LBSM research on geographic information can reflect the dynamic characteristics of urban vitality and crowd activities, and user information in metadata can reflect the spatial imagery of different groups. However, in previous studies, metadata based on LBSM and image data based on CV research are relatively independent, and many studies of geodata-based spatial analysis do not consider the image content of non-geodata images. Occupying a relative number of non-geodata images has great potential for mining, that’s why the combination of metadata and image data has research significance.

The image classification results for geodata and non-geodata images both show a strong public preference for skyscraper, bridge, promenade, and canal scenes. Geodata plays an important role in spatial analysis. Because only 18.9% of Flickr images having geographic coordinates, the distribution of these concentrations of the image of urban public space simply represents a high probability that the area has a more high-level public preference. Meanwhile, the non-geodata results show a preference for categories such as downtown, harbor, and water park, complementing geodata’s spatial analysis results and illustrating the role of non-geodata images in understanding images of urban public spaces. It is noteworthy that the weight of tourists in this finding is greater than that of residents. The analysis of locals’ imagery of urban public spaces remains valuable, as the proportion of the non-geodata images from locals remains high.

In response to the VGG-16 prediction results, the kernel density mapping results, and the current land use of the Haihe River, we propose three placemaking suggestions for urban public space in the Haihe River:
1. Enhance urban public space vitality strategy

The placemaking process can be used either in retrofitting an existing space or planning a new space. We suggest that the most valuable space for enhancing vitality is Area 1. To improve the vitality in Area 1, the existing space needs to be retrofitted, especially the Haihe Cultural Square, which should be the most dynamic area; currently, the area is not young enough, the urban vitality is seriously insufficient, and the landscape elements are relatively few. In addition, the square in front of the Tianjin Railway Station and the hydrophilic platform do not effectively combine commercial functions, and the diversity of urban public space in this area is low. A very good urban redevelopment is Osaka Station Plaza and the Umekita Plaza in Japan, which have been revitalized by combining the urban public space and the commercial center.

2. Urban public space reasonable diversity strategy

The Flickr prediction results also show relatively homogeneous landscape elements: that is, the center of gravity of people’s attention is skyscrapers, bridges, promenades, and urban/canals, and the images of urban public spaces that people pay more attention to are concentrated on the skyscrapers, bridge, and downtown in Area 1. We propose building more playgrounds and theme parks in Area 2 and developing more commercial pedestrian streets along the coastal streets in Area 3. In particular, Area 3 has greater potential and domestic demand due to the development of coastal residences.

3. Urban public space sustainable development strategy

From the perspective of improving the living environment of residents, urban public space regeneration in Area 3 is the most valuable; denser community parks and hydrophilic platforms can make this area a future community. From the perspective of tourism development, Area 1 and Area 2 have strong commercial potential and urban public space needs; this area can be planned for more consumer-oriented urban public space.

In this study, the main achievements, including contributions to the urban design field, can be summarized as follows:

1. This study demonstrates the outstanding contribution of social media image data as well as metadata to understanding images of urban public spaces while demonstrating the effectiveness of a regularized VGG-16 image classification model, Grad-CAM, and kernel density.

2. The results demonstrate the importance of Flickr geodata for the spatial analysis of urban public spaces and also demonstrate the limitations of studying images of urban public spaces only with results from kernel density. These findings are consistent with research showing that top-1 to top-5 image classification results for non-geodata images are particularly important.

3. This research method contributes several placemaking suggestions for urban public space regeneration around the Haihe River in Tianjin. This experimental research will hopefully serve as useful feedback for improving the vitality and diversity of urban public spaces.

6. Conclusions

The present results confirm that our optimized VGG-16 architecture is significant for Flickr image scene classification, including 81.75% top-1 accuracy and 96.75% top-5 accuracy, and is an effective Grad-CAM visualization method. This is an effective alternative to manual classification when the image data are complex and hard to distinguish. On this basis, we conclude that the overall images of urban public spaces are dominated by skyscrapers, bridges, promenades, urban canals, harbors, and parks on the Haihe River.

Three public spaces show concentrated image activity based on Flickr data with geographical coordinates: the area centered on Haihe Culture Square, the area centered on Tientsin Eye, and the river along Haihe E Road. The weights of these results are, respectively, 5% for residents, 15% for domestic tourists, and 80% for overseas tourists. However, it is difficult to supplement geo-information with non-geodata Flickr images.
by identifying image content. Despite these limitations, these results are valuable for classifying non-geographic images and spatially distributing images with geodata. The findings of this study can be understood as a criticism of using only images with geodata as research data. Images without geodata should be given more attention.

There are several limitations in the research methods and results of this study:

1. Incomplete Flickr metadata, including incomplete user information (age and gender) and incomplete geographic data. In this study, the data with geographic coordinates only account for 18.9% of the total data, and there were deviations in the accuracy of geographic coordinates in some specific circumstances. Such deviations are not prominent on the scale of urban planning or region planning, but on the scale of urban design, they seriously affect mapping results. Compared with Flickr users in North American cities, the number of Flickr users in Chinese cities is also insufficient. Furthermore, the geo-location of images is unable to accurately represent the locations of the public’s subjective preferences. For example, the public may stand opposite to shoot their favorite public space for better visual relations. That leads to the geographic coordinates of the taken photo location being naturally inaccurate. We believe that a research method that only uses metadata like geographic information or user information as the main data source is not suitable for Haihe River.

2. The image classification accuracy of VGG-16 is lower than that of new frameworks such as ResNet-101 (top-5 accuracy 95.6%) [67] or CoCa [68] (top-1 accuracy 91.0%). However, improving the training accuracy is not the future development direction of this research in the urban design field. The key to solving the problem is to interpret the deep learning results and use them to guide placemaking. For example, the river area in Haihe is a historical heritage protection district. There are not only many Italian-style, English-style, and French-style historical buildings but also modern buildings that were newly built while ensuring a unified architectural form. This is reflected in the fact that many Western-style villas are listed as churches, museums, and other buildings. There are also deviations in image classification that cannot be accurately identified by deep learning methods. The combination with architectural typology is also necessary.

3. The Places365 database has 1.8 million training images. The size of the training volume directly determines the prediction accuracy, but the disadvantage is that the training time is long and the requirement for graphics memory is high. If the results can be simplified to landscape and urban place-related datasets, it will greatly save time in model training and model optimization.

Future research on images of urban public spaces might extend the explanations of the relationship between people’s preferences for public space scenes and people’s emotions about these spaces. In summary, this paper argued that image classification based on the CNN network can provide an effective analysis of images of urban public spaces and makes suggestions for public space placemaking or regeneration. Overall, our results demonstrate the effects of computer vision techniques in urban design.

Author Contributions: Conceptualization, Chenghao Yang and Tongtong Liu; Data curation, Shengtian Zhang; Methodology, Chenghao Yang and Shengtian Zhang; Software, Chenghao Yang and Shengtian Zhang; Validation, Shengtian Zhang; Formal analysis and investigation, Chenghao Yang; Writing—original draft preparation, Chenghao Yang and Shengtian Zhang; Writing—review and editing, Tongtong Liu; Supervision, Tongtong Liu; Project administration, Tongtong Liu; Funding acquisition, Tongtong Liu and Chenghao Yang. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge funding from the China Scholarship Council (CSC) (Project NO.202106250093).

Data Availability Statement: Data are not available due to privacy restrictions.

Conflicts of Interest: The authors declare no conflict of interest.
32. Payntar, N.D.; Hsiao, W.-L.; Covey, R.A.; Grauman, K. Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. Tour. Manag. 2021, 82, 104165. [CrossRef]
33. Jingling, L.; Yun, L.; Liya, S.; Zhiguo, C.; Baqiang, Z. Public participation in water resources management of Haihe river basin, China: The analysis and evaluation of status quo. Procedia Environ. Sci. 2010, 2, 1750–1758. [CrossRef]
34. Shen, Y.; Wu, Z. Functional visibility graph analysis: Quantifying visuofunctional space with social media check-in data. Environ. Plan. B Urban Anal. City Sci. 2022, 49, 41–57. [CrossRef]
35. Konold, G.M.; Pinto, P.J. The social connectivity of urban rivers. Geomorphology 2017, 277, 182–196. [CrossRef]
36. Villani, C.; Talamini, G. Pedestrianised streets in the global neoliberal city: A battleground between hegemonic strategies of commodification and informal tactics of commoning. Cities 2021, 108, 102983. [CrossRef]
37. Chen, Y.; Liu, T.; Liu, W. Increasing the use of large-scale public open spaces: A case study of the North Central Axis Square in Shenzhen, China. Habitat Int. 2016, 53, 66–77. [CrossRef]
38. Jing, C.; Dong, M.; Du, M.; Zhu, Y.; Fu, J. Fine-Grained Spatiotemporal Dynamics of Inbound Tourists Based on Geotagged Photos: A Case Study in Beijing. IEEE Access 2020, 8, 28735–28745. [CrossRef]
39. 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. 2018, 40, 1452–1464. [CrossRef]
40. Cardoso, A.S.; Renna, F.; Moreno-Llorca, R.; Alcaraz-Segura, D.; Tabik, S.; Ladle, R.J.; Vaz, A.S. Classifying the content of social media images to support cultural ecosystem service assessments using deep learning models. Ecosyst. Serv. 2022, 54, 101410. [CrossRef]
41. Ding, X.; Fan, H. Exploring the Distribution Patterns of Flickr Photos. ISPRS Int. J. Geo-Inf. 2019, 8, 418. [CrossRef]
42. Vaisanen, T.; Heikinheimo, V.; Hippala, T.; Toivonen, T. Exploring human-nature interactions in parks with social media photographs and computer vision. Conser. Biol. 2021, 35, 424–436. [CrossRef] [PubMed]
43. Cao, R.; Zhu, J.; Tu, W.; Li, Q.; Cao, J.; Liu, B.; Zhang, Q.; Qiu, G. Integrating Aerial and Street View Images for Urban Land Use Classification. Remote Sens. 2018, 10, 1553. [CrossRef]
44. Chattopadhyay, A.; Sarkar, A.; Howlader, P.;Balasubramanian, V.N. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 839–847.
45. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 2015, 115, 211–225. [CrossRef] [PubMed]
46. Law, S.; Seresinhe, C.I.; Shen, Y.; Gutierrez-Roig, M. Street-Frontage-Net: Urban image classification using deep convolutional neural networks. Int. J. Geogr. Inf. Sci. 2018, 34, 681–707. [CrossRef]
47. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. Pattern Recognit. 2018, 77, 354–377. [CrossRef]
48. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
49. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of Tricks for Image Classification with Convolutional Neural Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.
50. Shafaeey, M.A.; Salem, M.A.-M.; Ebeid, H.; Al-Berry, M.; Tolba, M.F. Comparison of CNNs for remote sensing scene classification. In Proceedings of the 2018 13th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 18–19 December 2018; pp. 27–32.
51. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 2016, 54, 6232–6251. [CrossRef]
52. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 2014, 15, 1929–1958.
53. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural netwo55rks. Commun. ACM 2017, 60, 84–90. [CrossRef]
54. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
55. Silverman, B.W. Density Estimation for Statistics and Data Analysis; Routledge: London, UK, 1986.
56. Silverman, B.W. Density Estimation for Statistics and Data Analysis; Chapman & Hall/CRC: Boca Raton, FL, USA, 1998; p. ix. 175p.
57. Lloyd, A.; Cheshire, J. Deriving retail centre locations and catchments from geo-tagged Twitter data. Comput. Environ. Urban Syst. 2017, 61, 108–118. [CrossRef]
58. Huang, J.; Obracht-Prondzynska, H.; Kamrowska-Zaluska, D.; Sun, Y.; Li, L. The image of the City on social media: A comparative study using “Big Data” and “Small Data” methods in the Tri-City Region in Poland. Landsc. Urban Plan. 2021, 206, 103977. [CrossRef]
59. Villani, C.; Talamini, G. Pedestrianised streets in the global neoliberal city: A battleground between hegemonic strategies of commodification and informal tactics of conditioning. Cities 2021, 108, 102983. [CrossRef]
60. Chen, Y.; Liu, T.; Liu, W. Increasing the use of large-scale public open spaces: A case study of the North Central Axis Square in Shenzhen, China. Habitat Int. 2016, 53, 66–77. [CrossRef]
61. Silverman, B.W. Density Estimation for Statistics and Data Analysis; Routledge: London, UK, 2018.
62. Wu, J.; Chen, X.; Chen, S. Temporal Characteristics of Waterfronts in Wuhan City and People’s Behavioral Preferences Based on Social Media Data. *Sustainability* **2019**, *11*, 6308. [CrossRef]

63. Liu, K.; Moon, S. Dynamic Parallel Pyramid Networks for Scene Recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *1*–*11*. [CrossRef] [PubMed]

64. Qian, S.; Liu, H.; Liu, C.; Wu, S.; San Wong, H. Adaptive activation functions in convolutional neural networks. *Neurocomputing* **2018**, *272*, 204–212. [CrossRef]

65. Rizwan, M.; Wan, W. Big Data Analysis to Observe Check-in Behavior Using Location-Based Social Media Data. *Information* **2018**, *9*, 257. [CrossRef]

66. Słomska-Przech, K.; Gołębiowska, I.M. Do different map types support map reading equally? Comparing choropleth, graduated symbols, and isoline maps for map use tasks. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 69. [CrossRef]

67. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.

68. Wortsman, M.; Ilharco, G.; Gadre, S.Y.; Roelofs, R.; Gontijo-Lopes, R.; Morcos, A.S.; Namkoong, H.; Farhadi, A.; Carmon, Y.; Kornblith, S. Model soups: Averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. *arXiv* **2022**, arXiv:2203.05482.