This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3395469

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.
Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Behavior Analysis of Photo-Taking Tourists at Attraction-level Using Deep Learning and Latent Dirichlet Allocation in Conjunction with Kernel Density Estimation

Muhammad Iqbal [1], Li Renjie, and Jingyi Xu
[1]School of Geographical Sciences, Hebei Normal University, Shijiazhuang, People's Republic of China

Corresponding author: Muhammad Iqbal (usafzai@hebtu.edu.cn).

ABSTRACT User-generated content (UGC) on social media platforms plays a significant role in conveying individual sentiments that are effective in predicting image sentiments by evaluating its contents to excavate the behaviour and cognition of image producers at attraction-level tourist destinations. In this study, we aim at the attraction-level study of the spatiotemporal behavior of photo-taking tourists. We used the deep convolutional neural network model DeepSentiBank (DSB) for sentiment prediction of Flickr photos. Then, Latent Dirichlet Allocation (LDA) was employed to categorize these sentiment predictions according to the noun content of specified distinct groups. Next, Kernel Density Estimation (KDE) was used as a spatial analysis tool to determine the spatial distribution characteristics of tourists' behaviors. The photo-taking behavioural patterns of different tourist types are analyzed from the individual preferences of tourists and depicted in terms of both visual semantics and spatiotemporal behavioral features. It is observed that the proportion of landscape preferences of the thematic tourists are not significantly affected by seasons but by the aggregate activities of thematic tourists that significantly vary in autumn. The findings obtained from the analysis hold immense significance in the realm of tourism and hospitality management which can play a significant role in the development of facility management, regional tourism, and prospects.

INDEX TERMS Adjective Noun Pairs; Attraction Level Activities; Deep Learning; Kernel Density Estimation; La-tent Dirichlet Allocation; Spatiotemporal Behavior

I. INTRODUCTION

Photography is a favorite activity of tourists because it records their fascinating memories of the surroundings [1]. The global establishment of a sustainable industry for tourism needs an understanding of the tourism phenomenon and the travel behavior of tourists. Energy and capital consuming data collection by conventional means usually distort the findings. However, the popularity of social media skyrocketing generated a research gap on visiting and travelling patterns using geotagged images acquired from social media platforms [2], [3]. Enhancement of advanced web technologies as well as an increase in the availability of devices equipped with Global Positioning System (GPS) enabled ordinary people to produce spatial data [4]. This spatial data or UGC is in the form of blogs [5], travelogues [6], and images [7] posted to national and international directories like Weibo, Go.com, Flickr, TikTok, Instagram, Facebook, Twitter, etc. Significant UGC material in the form of text and photographs containing rich attribute data sometimes in the form of Exchangeable Image Files (EX-IF) offers a rich data foundation for tourism-related research. Therefore, the study of tourist travel patterns using data from web photo databases with geotagged information has real-world applications [8], [9]. This information can be used to analyze the spatiotemporal behavioral characteristics of tourists [10], [11], and destination resource allocation [12].

The spatiotemporal details of tourists' behavior help to understand travel activities [13]. These activities are often performed at three scales [14] namely, destination-to-

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3395469

destination, destination-level, and attraction-level [15]. Destination-to-destination describes the behavioral movement from exit to destination or between two destinations, such as countries [16], regions or cities [17]. The destination-level tourist behavior focuses on the trajectory of visitors' movements within a location, such as the analysis of visitors' behavior in various tourist attractions in a city [18], [19]. On the contrary, the attraction- level study of tourists' behavior targets the patterns of tourist movements within the attraction [14]. The investigation of tourists' spatiotemporal behavior at a microscopic scale pertains specifically to the movement occurring within an attraction due to its localized nature [20]. Attraction-level tourist movement takes place within a confined area of limited dimensions and a well-defined perimeter as exemplified by theme parks [21], [22], where tourist behavior can be subjected to greater regulation. Considerable focus has been directed towards exploring destination-to-destination and destination-level behavior of tourists. However, researchers have paid relatively little attention towards scrutinizing the attraction-level behavior of visitors.

During their journey, visitors undergo diverse emotional encounters, encompassing emotions like joy, dissatisfaction, enthusiasm, frustration, and remorse, among others [23]. Factors attributing to tourist discontent predominantly comprise unfavorable service dispositions, adverse weather conditions, overcrowding at attractions, unexpected calamities, and so forth. Forecasting too many sentiment-related particulars conveyed by images taken by visitors involve the utilization of SentiBank, a collection equipped with detectors for Adjective-Noun Pairs (ANPs) [24]. This resource facilitates the examination of image sentiments via analytical constructs linked with ANPs. In the context of training the ANP sentiment dataset, which encompasses numerous categories and incorporates over a million images sourced from Flickr, it is noteworthy that recent research [25] has demonstrated remarkable enhancements achieved by deep convolutional neural networks in terms of classification accuracy and efficiency, particularly on datasets comparable to ImageNet. These neural networks show a considerable learning capacity, adaptable to various network depths and widths. Additionally, their robust modelling of statistical stationarity and the inherent pixel dependencies tied to image nature is notably accurate. The advantages of convolutional neural networks extend to their ease of training, stemming from consistent layer sizes. This architectural choice results in a lower number of connections and parameters, with only minimal impact on theoretical performance. Additionally, these networks possess the capability to leverage model weights acquired from more comprehensive datasets. This adaptability allows models initially trained on datasets like ImageNet to be effectively applied in specific scenarios. For instance, in this study, the model learned from ImageNet is effectively transferred to a specialized dataset like SentiBank [26].

Conventional methods such as questionnaires and time-space diaries can gather data on various aspects of tourists' behavior such as personal traits, travel expenditures, and travel patterns [14]. Such information holds significance in analyzing the fundamental characteristics of tourist behavior. Time-space diaries are particularly believed to provide comprehensive insights into behavioral patterns [27]. However, gathering data in a manner that necessitates the documentation of travel history by participants can significantly skew the results. The conventional data collection methods are often restrictive, impeding the exploration of tourist' conduct. Conversely, the advent of location-based technology made it feasible to employ UGC trajectories for monitoring tourist movements. These trajectories have been utilized to examine the spatiotemporal patterns of tourists' movement behavior at the microscopic level in Mount Huashan [14], Akko Old City [28], Giant Panda National Park, China [27], and Lion Grove [29]. The utilization of GPS data in microscopic scale research has gained significant attention, allowing for the investigation of spatiotemporal patterns of tourists' movement behavior at the attraction-level.

Mining the visual content of tourism photos primarily relies on findings stemming from manual image recognition in early relevant investigations. For instance, Hunter's work involved a content analysis technique to examine web images of Seoul, Korea [30]. Similarly, Zheng adopted the NViov10 qualitative coding method to construct visual representations of tourism location images [31]. In separate studies, Xingzhu, Kai [32] and Jing, Xingzhu [33] employed geo-tagged photos to scrutinize spatial attributes of tourism flows. Meanwhile, Kuo, Chan [34] leveraged geographical data from location photos in tandem with spatial analysis approaches for exploring Point-Of-Interest (POI) and Area-Of-Interest (AOI) identification. Mou, Yuan [35] investigated variations in urban inbound tourism patterns using similar techniques. Dunning, Lina [7] merged textual information from photo metadata, traditionally seen as indicative of tourist perceptions, with visual studies. They employed comment labels from picture metadata to map cognitive and emotional imagery to analyze perceptions of tourist sites. The evolution of computer technology paved the way for machine recognition of extensive image datasets, ushering in tourism-related studies grounded in deep learning photo recognition. Zhang, Chen [36] performed scene recognition on tourist photos taken in Beijing, comparing behavioral and cognitive distinctions among tourists, irrespective of their origins. Deng, Liu's [37] research relied on deep learning analysis of images to probe distinct perceptions of tourist destinations. Yuehao, Ying [38] carried out a comparative investigation into the imagery associated with 24 major Chinese cities, utilizing machine-tagged text from Flickr photo metadata. Bubalo, van Zanten [39] amassed geographical data concerning landscape preferences and perceptions through various crowdsourcing models.

Landscape perception encompasses the intricate interplay between individuals and their surroundings. It involves the formation of perceptions through experiential activities and the subsequent impact of these perceptions on both individuals and the landscape itself. The scope of landscape perception encompasses various aspects, including types of landscape perception, preferences, attitudes, and the perception of the value associated with landscape [23]. An increasingly promising avenue of research lies in utilizing visual content from photographs to analyze tourist' perceptions of landscape and their emotional preferences when visiting tourism destinations. Dunkel [40], for instance, leveraged Flickr photo data to introduce a comprehensive method for calculating landscape perceptions. Similarly, Figueroa-Alfaro et al. [41] employed location-based photo data to assess Nebraska's landscape, identifying hotspot areas and aesthetic values. The focus of landscape perception research has evolved from mere examinations of the visual landscape itself to investigations into its connections with other tourism-related variables. Wang and colleagues [42] elucidated the correlation between visual landscape presentation and the overall tourism experience within ancient villages by categorizing the visual landscape. Meanwhile, Zheng et al. [43] delved into the relationship between landscape perception and other variables, considering factors like tourist satisfaction and perception. Deep learning and big data mining technologies continue to advance rapidly. Likewise, computer image processing technology is also evolving and becoming increasingly sophisticated [44]. Consequently, various fields are adopting Artificial Intelligence (AI) for image recognition purposes [45]. The application of AI-based big data processing techniques in the analysis and recognition of location photographs captured by tourists overcomes the constraints of manual approaches and offers technological assistance in extracting intricate visual and semantic data from such images [46].

This paper investigates the landscape perception-based types of tourists by utilizing technology for location-based photo visual semantic mining at the attraction-level. The study employs visual semantics extracted from tourists' location photos to classify tourist types based on captured landscape perceptions, and preferences of tourists at a microscopic spatiotemporal scale. We used deep learning to analyze the visual semantics of tourist photos taken in the Old Town of Lijiang. The location and time information embedded in the photos are used to identify landscape perception types, the spatiotemporal trajectory of thematic tourists and the visual sentiment preferences of tourists' photos at a temporal (seasonal) scale. Furthermore, this research explores various spatial scales, such as Historical and Cultural (T1) (Architecture, Religion, Monuments, Chinese Characters), Human Life (T2) (People, Pets, Food, Activities), Natural Environment (T3) (Landscape, Meteorology, Environment), and Urban Spatial (T4) (Neighborhoods, Public Spaces, Transportation, Infrastructure) on which the types of tourists are classified showing different sentiment preferences.

Our primary contributions include:

1. We utilize the deep learning technique DSB to calculate the top 5 ANPs demonstrating the top five visual feature contents with the highest confidence.
2. We categorize the photos from the noun content into distinct groups using LDA.
3. We employ KDE as a spatial analysis tool to determine the spatial distribution characteristics of tourists' behaviors.
4. We classify the tourists based on the captured theme.
5. We provide extensive analysis of photo-taking tourists to understand their spatiotemporal behavior at the attraction-level.
6. We analyze the photo-taking behavioral patterns of different tourist types from individual preferences and depict them in terms of both visual semantics and spatiotemporal behavioral features.

The rest of the article is structured as follows: First, Section II describes the study area, explains the data collection process, and presents the proposed methodology. Next, Section III provides the results of the proposed methodology. Then, Section IV offers a detailed discussion of the obtained results. Section V highlights the limitations of the proposed approach. Section VI concludes this article at the end.

II. MATERIALS AND METHODS

A. Study Area

The Old Town of Lijiang (26.8718º N, 100.2359º E) in Yunnan Province, China encompasses three distinct areas: Dayan old town, Basha housing cluster, and Shuhe housing cluster. The term "Old town of Lijiang" refers specifically to Dayan's old town. The Old Town of Lijiang is approximately 1.6 km^2 (Figure 1) and is surrounded by the Lion Mountain to the west and the Elephant and Golden Row Mountains to the north. The Old Town has a rich history dating back to the Song and Ming dynasties and has maintained its original, high-quality architecture [48]. The buildings are notable for their fusion of cultural elements from various cultures that have merged over time. It is renowned for its well-organized system of waterways and bridges. The Black Dragon Pool, located at the foot of Elephant Mountain, serves as the primary water source for the Old Town. The pool water flows from north to south,

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

split into three streams by the Shuangshi Bridge, and further divided into countless smaller streams that supply water to each household. The complex network of streams comprises about 354 bridges, making it the area with the highest bridge density in China. Additionally, various water wells and small ponds are distributed throughout the Old Town with Sanyanjing being the most well-known, consisting of three ponds with different functions. This Town was officially listed on the United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage List on December 4, 1997.

This ancient town has garnered worldwide recognition and established itself as a highly regarded destination in the competitive tourism industry. The old town welcomed approximately 0.22 million visitors in 1994. However, following its inclusion in the World Heritage List, the number of tourists significantly increased to 1.7 million in 1997, 9.1 million in 2010, and 54 million in 2019. A rise of 7.99 per cent in the local revenue generated by the tourism industry was observed in 2019, amounting to 107.8 billion RMB (equivalent to approximately 16.7 billion US dollars) when compared to the preceding year [47]. Our study primarily stems from the availability of data, which was accessible up to 2014, and regrettably, no future data was obtainable. We emphasized the significance of qualitative depth in comprehending tourists' perceptions and emotions, wherein a smaller but representative sample often proves more effective in delivering rich insights.

B. Data Source and Processing

The information used for this research is drawn from a set of 100 million publicly available multimedia from the Yahoo Flickr database. This Yahoo Flickr Creative Common 100 million Dataset or YFCC 100M dataset [48] is the Yahoo Web Scope Project part. The metadata from Flickr (2004-2014) was downloaded with the keyword "Lijiang", which was later precisely cropped in ArcGIS software using the administrative boundary of "Lijiang County". The data was further cropped to our study area after processing.

To describe the spatiotemporal behaviour of tourists, locals and tourists with multiple visits have been isolated. Moreover, due to improper positioning of photography equipment or other factors, there is an overlap in the location information of images shared by some users, which can minimize the results of our interpretation of tourists. To improve the accuracy and maintain the scientific nature of the study, we systematically assume that the stay of a photographer for more than 10 days will be considered as a local or multiple-visit, visitor in whole Lijiang County. Moreover, three or fewer shots from the camera with the same location information will not be considered a tourist. The assumption about tourists can be seen on the scatter chart (Fig. A1).

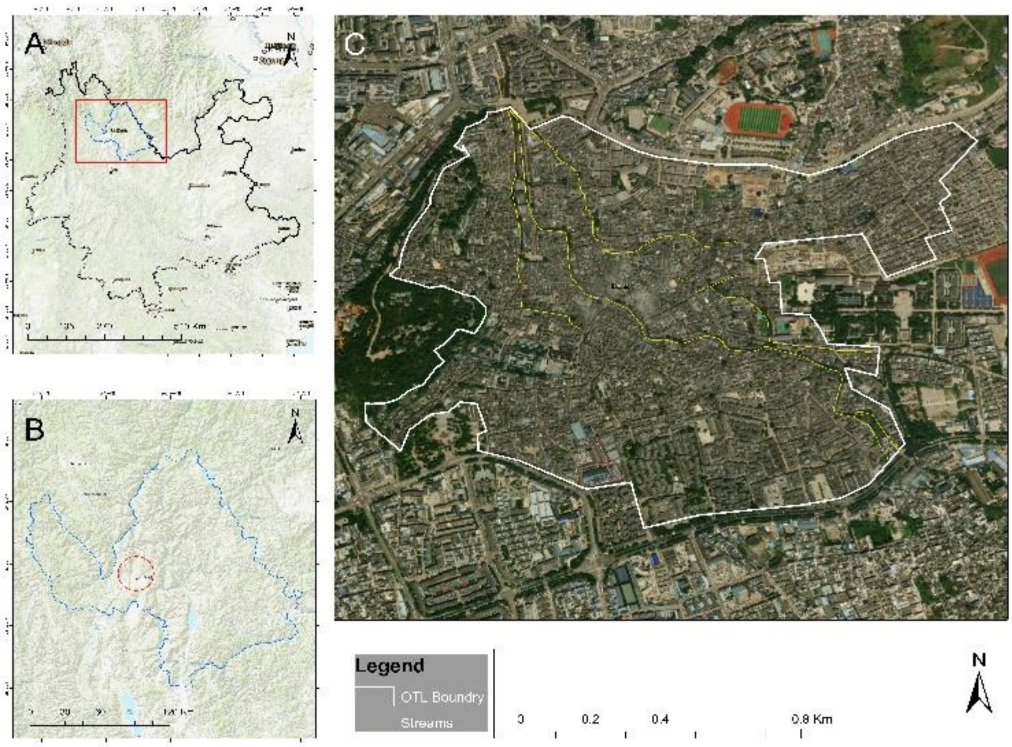

Figure 1. **(A) Lijiang County (B) Study Area (C) and Position in Yunnan Province**

Most of the tourists are concentrated in less than 10 days' span. The appropriate period is stated as 10 days or fewer to ensure the amount of study. It is also specified that the number of images taken by the same tourist with varied location information must be larger than or equal to three, which is deemed to have spatial information validity. All tourists of Lijiang County were calculated and screened using the above period and spatial information validity, and then the "Old Town of Lijiang" (OTL) tourists were cropped for further study. Out of 4545 photos and 196 tourists (Lijiang County), 1537 photos and 53 tourists (Old Town Lijiang) were identified to be the representative sample tourists.

C. Proposed Methodology

The proposed methodology is illustrated in Figure 2. In step 1, tourism data about Lijiang County was collected and was further screened for further processing. In step 2, a deep learning model is used to recognize the contents of tourists' captured images. The objective of photo recognition is to categorize tourist images according to their noun content where each photo would be a member of a distinct group determined by the LDA topic extraction model. In step 3, tourists are classified based on captured themes, where spatial analysis is performed to determine the spatial distribution characteristics of tourists' behavior.

The existing literature frequently used one highest confidence recognition result to determine the content of an image; however, tourist-taken photos frequently demonstrate multiple visual contents. Therefore, the highest confidence recognition results might not be able to accurately describe the subject imagery of photographs.

Please note that the experiments are performed using 64-bit operating system with the following specifications.

- Intel i5 11th Generation with 2.40 GHz processor
- 16 GB RAM

Similarly, the software used is ArcGIS with following specifications.

- ArcGIS Desktop version 10.8.2
- License Type: Advanced

Each of the steps in the proposed methodology is explained in detail in the following sections.

1) Convolutional Neural Networks for Visual Classification Model

The DSB approach utilizes psychological principles and web mining techniques to create a visual sentiment ontology, a vast detector library, and a visual sentiment test benchmark. Its concept classifier was trained on more than a million geotagged photos, which enabled it to produce 2089 ANPs comprising 231 adjectives and 424 nouns. This process converts image data into textual information. DSB is based on the deep learning framework where trained DSB greatly increases the annotation accuracy in ANPs based on the performance evaluation and comparisons with its predecessors [26]. In this article, the visual content classification results of each photograph are based on the top five ANPs in each photo's confidence ranking. Based on this content, classifications of tourists are innovatively identified on their preferred landscape capture.

We harnessed the concept of transfer learning, a well-established technique in the realm of deep learning. Sentibank, being a pre-trained model, arrives with a wealth of knowledge gleaned from its training on extensive datasets.

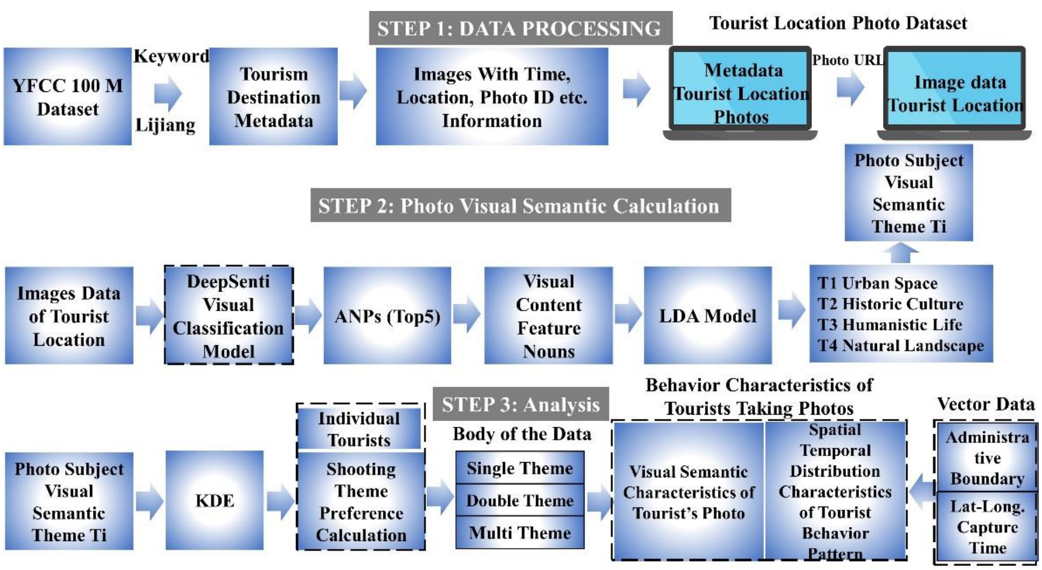

Figure 2. **Flowchart Illustrating the Research Process**

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

This reservoir of prior knowledge substantially facilitates the model's adaptability to our specific sentiment analysis task, even when confronted with relatively smaller datasets. Our meticulous screening process was instrumental in ensuring that the retained dataset exhibited high quality and aptly represented the diverse spectrum of tourist experiences within our study area. This targeted approach allowed us to extract profound and meaningful insights from the data.

Network Architecture

Figure 3 illustrates the proposed deep CNN architecture that follows the architecture of AlexNet [25]. This network is composed of eight blocks where the first five are the convolution blocks and the last three are fully connected blocks. The first two convolution blocks contain convolutional layers followed by ReLU (Rectifier linear unit) layer, Pooling layer, and local response normalization layer. Blocks 3, 4, and 5 involve convolutional layers followed by only the ReLU layer. In block 5, the max pooling layer is also used after the ReLU layer. Blocks 6 and 7 consist of fully connected layers followed by the ReLU layer and dropout layer. The last block of our network is composed of only fully connected layer that provides its 2089-dimensional output to the softmax activation function for the final classification of the visual sentiment concept.

The purpose is to maximize the log probability average of the training set. The second, fourth, and fifth layers of our proposed network connect their kernels to only half of the kernel maps of the previous layers whereas the third layer connects its kernels to all the kernels of the second layer. Note that non-linear ReLU is used in all convolutions and fully connected blocks except for the last one where the softmax activation function is used for the final classification. The max pooling operations are used only in the first, second, and fifth blocks and the dropout layers are only used in blocks 6 and 7.

The response normalized activity is computed using Equation 1.

$$b^{i}_{x,y} = a^{i}_{x,y} / \left(k + a \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+1/2)} \left(a^{i}_{x,y}\right)^2\right)^{\beta} \qquad (1)$$

Where $a^{i}_{x,y}$ is the output of the max pooling operation that indicates the neuron activity, the sum is obtained by combining n kernel maps at different positions, and N indicates the count of kernels in the layer. We used the following values for different variables: α = 10-4, β = 0.75, k = 2, and n = 5.

The dimensions of all the images in training and testing sets are set to 3 x 256 x 256 before submission to the network without the aspect ratio. We also applied image translation along with horizontal reflections to the data to avoid overfitting. This is achieved by randomly extracting 227 x 227 patches along with their horizontal reflections from the normalized input (i.e. 256 x 256). These patches were used to train the network. Therefore, the first convolutional layer in block 1 uses 96 kernels each of size 11 x 11 x 3 with a 4-pixels stride to filter the 227 x 227 x 3 patch. Similarly, the convolutional layer in block 2 receives the output of block 1 that is filtered using 256 kernels each of dimension 5 x 5 x 48. Likewise, convolutional layers in block 3 and block 4 use 384 kernels each of size 3 x 3 x 256 and 3 x 3 x 192, respectively. The convolutional layer in block 5 employs 256 kernels each of size 3 x 3 x 192. Each of the fully connected layers in blocks 6 and 7 has 4096 neurons.

Model Learning

In this work, we used stochastic gradient descent to minimize the regression objective with the following parameter values. Batch size was set to 256 samples whereas momentum and weight decay values were set to 0.9 and 0.0005, respectively. It is worth noting that the weight decay value helps in the regularization of the model as well as reduces the training error. Our network may demonstrate biased behavior to those images that have stronger sentiments, therefore, overfitting may be a problem. Similarly, small data may also contribute to the overfitting of our network.

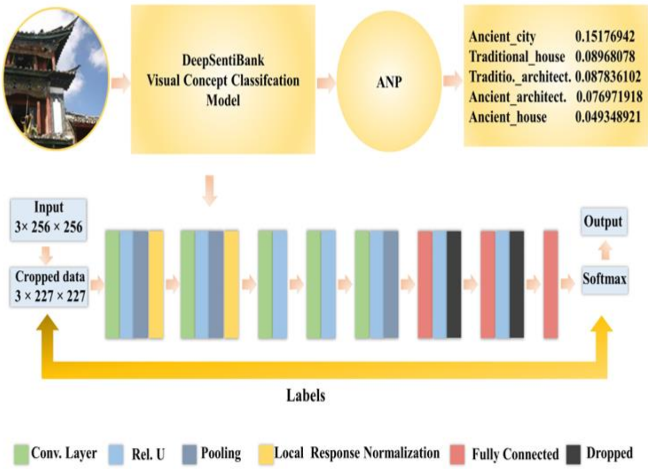

Figure 3. **DSB model with convolutional neural network structure**

Therefore, we used the pre-trained model of ILSVRC2012 for fine-tuning and hence initialized the weights of our network with those parameters excluding the top layer. The learning rate parameter is set to 0.001. The proposed network is trained for 77 epochs which equals roughly 250,000 iterations. The learning rate is reduced by a factor of 10 for every 20 epochs which equals 100,000 iterations.

2) Noun-based Classification using LDA

Visual content analysis is a rigorous, observational method for investigating how a phenomenon is represented [49]. It enables the quantification of observable content samples classified into distinct groups [50]. The primary objective of content analysis is to identify the key focal points in the pictures and code their frequencies, co-occurrences, clustering, and other concerned issues [46]. The effective fitting of models to big datasets is a major research subject in topic modelling [51], [52], [53]. To illustrate the perception differences among tourists, we use LDA, which is an auto-encoder model of text. The core idea is that the documents are represented as random mixtures of discriminative features with each subject defined by a word distribution [53]. All the nouns of identified photos in the ANPs are extracted as visual content feature words and categorized by LDA. The results of content feature words are high frequency-based nouns of the top 5 confidence ranking features. After the elimination of duplicate words, the visual features are finally nominalized in the following four categories (Table 1).

The categories T1, T2, T3, and T4 represent four perceptions and each perception covers some contents. Each category includes a number of visual elements. So, a photo will be categorized in single or multiple perceptions if it has these visual elements.

TABLE 1. NOUN-BASED CLASSIFICATION USING LDA

	Perception	Content	Included Visual Elements
T1	Historical and Cultural Landscape	Architecture, Religion, Monuments, Chinese Characters	History, Ice, Horse, Wonder, Paintings, Drawing, Cross, Skull, Science, Soldier, Head, Sand, Cemetery, Religion, Museum, God, Worship, Heritage, Armory, Bones, Arts, Zombie, Meditation, Pony, Lighthouse, Design, Sign, Architecture, Castle, Monument, Places, Fortress, Spider, Museum, Dragon, Graveyard, Landmark, Tower, King, Church, Metal., Worship, Prayer, Asylum, Grave, Statue, Sculpture, Face, Wall
T2	Human Life Landscape	People, Pets, Food, Activities	Lady, Food, Cat, Girls, Crowd, Kids, Dress, Student, Wedding, Cake, Dance, Beer, Child, Smile, Hair, Meat, Drink, Party, Race, Lights, Glass, Glasses, Chocolate, Baby, Sports, Children, Family, Apple, Puppy. Christmas, Egg, Worker, Celebration, Community, Cage, Business, Economy, Chair, Lego, Commercial, Fire, Cruise, Attack, Adventure, Festival, Holiday, Dog, Guy, Animals
T3	Natural Environment Landscape	Landscape, Meteorology, Environment	Spring, Landscape, Beach, Park, Garden, River, Tree, Trees, Lake, Night, View, Morning, Fog, Hills, Grass, Valley, Sunrise, Farm, Rain, Flower, Flowers, Leaves, Sunset, Plant, Rose, Blossom, Mountains, Sunlight, Flora, Butterfly
T4	Urban Spatial Landscape	Neighbourhoods, Public Spaces, Transportation, Infrastructure	Room, Office, Space, Airport, Hospital, Construction, Industry, Farm, Car, Bridge, Road, Signal, Street, Streets, Building, House, City, Hotel, Market, Home, Hall, Factory, Train, Theatre

3) Kernel Density Estimation Mapping

The application of KDE mapping is a useful tool to analyze the statistical outcomes and identify the hotspots of tourists' behaviors. KDE mapping has the capability of identifying hotspots visually from huge datasets [54]. Authors in [55] have used KDE for the exploration of Twitter activity. Formally, the KDE function [57] is defined in Equation 2.

$$f(x) = \sum_{i=1}^{N} \propto_i k_h(x - x_i) \qquad (2)$$

which returns the estimated density at a location x. The $\propto_i$ is the kernel weights with $\sum_{i=1}^{N} \propto_i = 1$. Usually, all the kernels are equally weighted, $\propto_i = 1/N$. The kernel function $k_h(x)$ is required to satisfy $\int k_h(x)\,dx = 1$ and $k_h(x) \geq 0 \;\forall \times \in R^2$. More important than the kernel function is the bandwidth h, which controls the degree of smoothing of the surface.

This method is used for non-parametric estimation of the probability density function of a random variable. The technique involves placing a kernel, usually a Gaussian kernel, at each data point and then summing the kernels to create a continuous estimate of the density function. The kernel bandwidth controls the smoothing of the data, determining the size of the region over which the data are smoothed. A larger bandwidth results in a smoother estimate whereas a smaller bandwidth produces a more jagged estimate. In this article in the case of circular region with a 100 square meter radius, the kernel density estimate is smoothed over a circular area of that size. The smoothing parameter is responsible for deciding how much weight is given to nearby data points, which impact the level of detail in the density function estimation.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3395469

III. RESULTS

A. The Visual Semantic Quantification of Different Thematic Landscapes

Figure 4 illustrates the visual semantic quantification of different thematic landscapes. In this regard, we took all photos as the analysis scale and the spatial kernel density is calculated for varied perceptions of tourists to visualize the spatial distribution characteristics of different thematic landscapes in OTL.

The kernel density image size is set to 5 and the search radius is 100 square meters based on the spatial scale of OTL. The spatial distribution of the four types of thematic images demonstrates a spatial characteristic pattern of dispersal from the hot core to the surrounding area with OTL's iconic and famous attractions. Generally, the most popular core-covered attractions with highly intensive density are the surroundings of Sakura Kim, Blue Papaya, and the Beaty Inn hotel in the historical and cultural landscape (T1). The Human Life Landscape (T2) hot core area covers the sides of Tibet Café, Old Town Spring Supermarket, Mei Shi Guang Chang, and Lijiang Lize Graceland Merry Inn. Natural environment (T3) shows hotspots with a high density near Lijiang Lize Graceland Merry Inn, Old Town Supermarket and Blue Papaya. The Beaty Inn and Vsherry Restaurant are also surrounded by an intensive core of density. Urban spatial landscape (T4) hot cores cover Yuhezhong River in the centre and at Beaty Inn and Lijiang Lize Graceland Merry Inn.

Within our dataset, comprising 1537 photos and involving 53 tourists, it is worth noting that the participation of tourists in each thematic landscape is nearly uniform. The T1 theme was captured by 34 tourists, T2 by 38, T3 by 40 and T4 theme was captured by 38 tourists. This observation reveals a recurring overlap in the spatiotemporal trajectories of these tourists. Their revisits to the same locations indicate the magnetic allure of these places. However, it is essential to acknowledge that similar geographic areas can lead to overlapping thematic interests among tourists, a phenomenon frequently encountered in our study.

B. Patterns of Photo taking Behavior of tourists in the Old Town of Lijiang

To explore the characteristics of different types of tourists and their behavioral patterns, we organize the visual semantic quantification results in 15 possible combinations based on the tourists as photographers in (Table 2) which is mathematically expressed as:

$$2^4 - 1$$

Where "4" is the number of themes of visual semantics and "1" is the empty set.

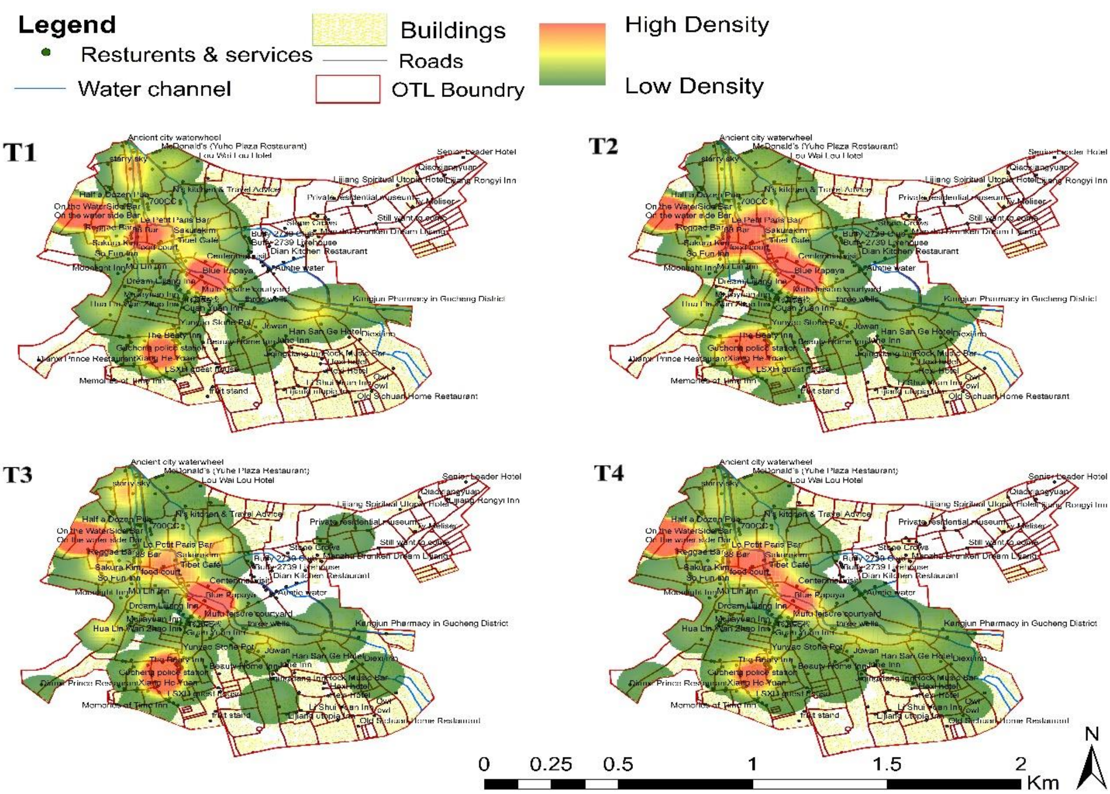

Figure 4. Visual Semantic Quantification of Different Thematic Landscapes

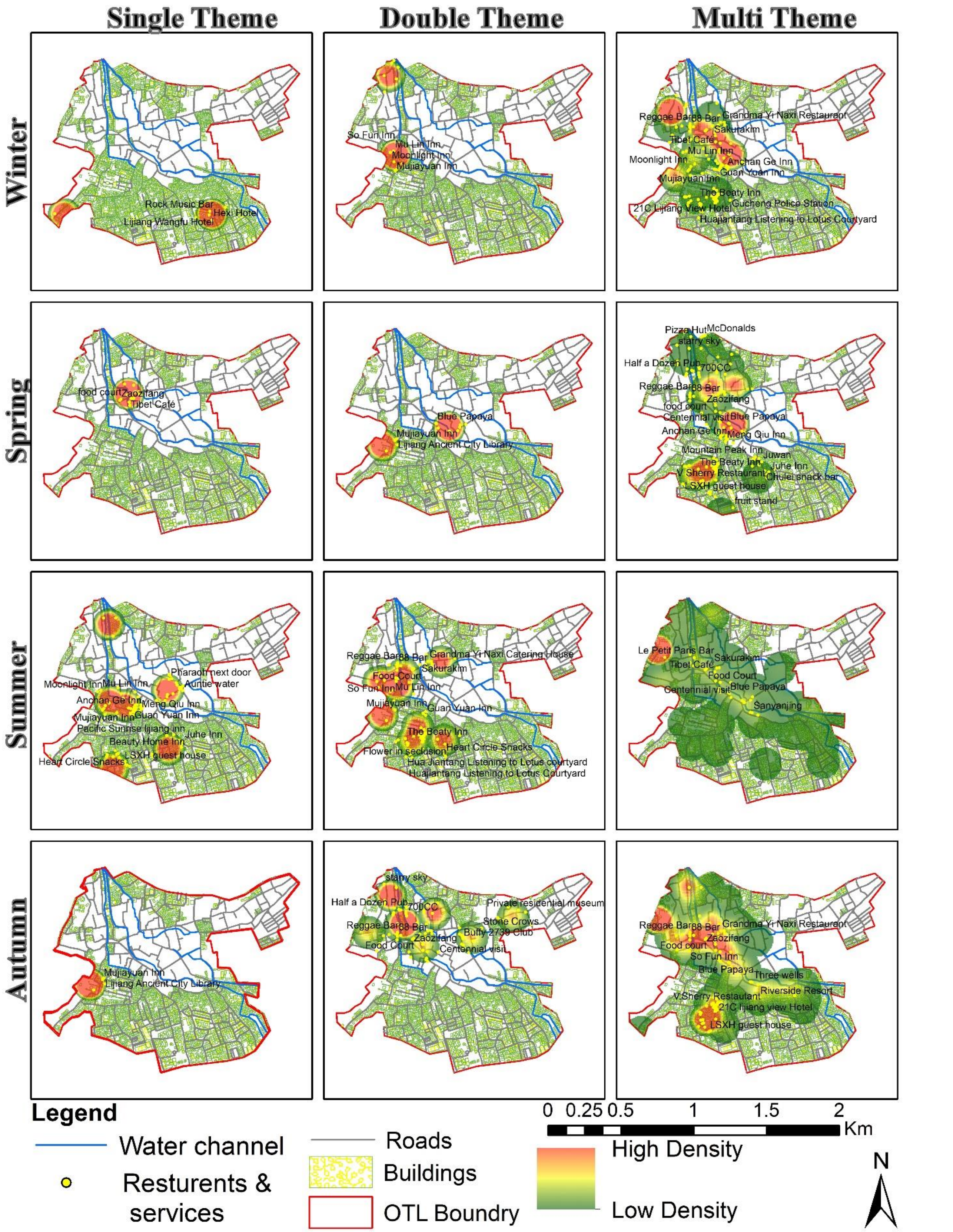

Figure 5. Seasonal Behavior of Tourists in the Old Town of Lijiang

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3395469

Single theme capturing tourists make the lowest proportion of photographs which is about 1% of the total photos. Overall, 18.8% of the tourists belong to a single theme and have a strong purpose and interest in shooting photos during their trip. Most of their photos have the same visual elements and are taken in the same style and angle.

Relative to single theme, dual theme photographers tend to capture a high proportion of photos with two visual themes making 2.7% of the total photos with 22.6% tourists involved. They show interest in capturing well-defined themes.

TABLE 2. PATTERNS OF PHOTO TAKING BEHAVIOR TOURISTS IN OTL

Tourist Type	Subordinate Landscape	Tourists Proportion		Photo's Proportion
Single theme	T1	1	18.8%	1%
	T2	4		
	T3	2		
	T4	3		
Double/Dual theme	T1T2	1	22.6%	2.7%
	T1T3	2		
	T1T4	1		
	T2T3	2		
	T2T4	2		
	T3T4	4		
Multi-theme	T1T2T3	3	58.5%	96%
	T1T2T4	2		
	T1T3T4	1		
	T2T3T4	2		
	T1T2T3T4	23		

Multi-theme capturing tourists are the most numerous of all types responsible for 96% of the total photos with 58.5% of photographers involved. The majority of them covered all the themes in their photos. This group of tourists has more comprehensive observation and rich experience towards the touristic places. They do not limit themselves to specific landscapes rather capturing all concerned places. Multi-theme capturing tourists involve a wide range of irregular photo shoots, enhancing the diversity in the visual semantic content of the images.

C. Patterns of Spatiotemporal (Seasonal) Behavior of Tourists in Old Town of Lijiang

To emphasize specific points of interest, we opted for partial labelling on the maps to maintain visual clarity. Seasonal interest in shooting different themes of the Old Town of Lijiang was portrayed in (Figure 5). The most populated season as far as the number of photos is concerned is autumn, which received 740 photos captured by 14 photographers. It is observed that individuals who showed greater enthusiasm for photography tended to capture a larger number of photos and visit a higher number of attractions in comparison to those who were less interested. Out of these 740 photos, only 6 photos are a single theme, 18 double themes, and 716 are multi-theme captured by two, four, and eight photographers, respectively.

Autumn is followed by Spring as the season of attraction for tourists received 297 photos by 12 photographers. Out of these, 297 photos were multi-theme, 10 dual themes, and only one photo was single theme captured by nine, two and one photographers, respectively. Summer received the third position in photo capture and stood 1st among several visitors. Largely, 285 photos were clicked by 16 number of visitors.

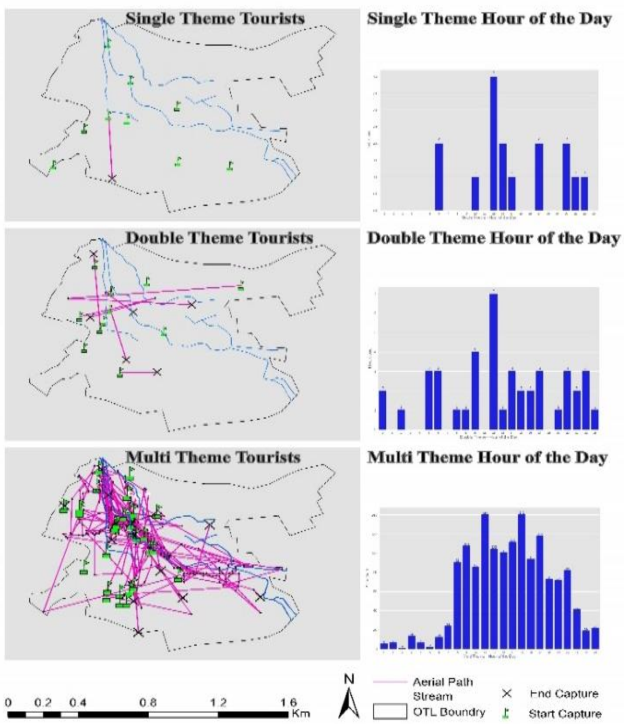

Figure 6. Photo proportion patterns of seasonal preferences of thematic tourists

Among these visitors, six showed interest in becoming multi-theme visitors by capturing 268 photos. Some five photographers took 11 dual theme photos and single theme five photographers captured only six photos. The least number of photographs were experienced in winter where the total number of photos is 215 by 11 photographers, 208 photos were multi-theme by eight photographers, dual theme has four photos by one, and single theme has three photos by one photographer. Monthly and hourly tourist activities can also be understood by the clock diagram (Fig. A2). Similarly, the proportion of seasonal behavior of tourists in the Old Town of Lijiang, patterns of temporal landscape preferences of thematic tourists across different seasons can be explained (Figure 6).

D. Pattern Analysis of the Spatial-Temporal Trajectory of Tourists in the Old Town of Lijiang

To elaborate on the spatiotemporal information of tourists, we used ArcGIS to discover the starting (green flag) and ending (black cross) points (Figure 7) of different categorical tourist's attention. We also visualized the cluster trajectory and the average time spent by different types of tourists during their photo shoots. The overall trajectory of photo-taking behavior of single-theme, double theme and multi-theme visitors is portrayed. The route pattern of the single theme tourists is very simple involving fewer POIs or considering only a single photo to leave a footprint. Single theme tourists show a scattered starting photoshoot point (green flags) as displayed in (Figure 7) that has a specific capturing aim of the single theme seekers.

Table 3 Average value of stay and photos per tourist in OTL.

Average Value	Single theme	Double theme	Multi-theme
Stay in Hour	8.4	23.3	66
Photos/tourist	1.6	3.6	47.7

They show interest in photo taking continually during the daytime from 12:00 PM to 02:00 PM and evening time 8:00 PM to 10:00 PM. The average stay of single theme tourists during their photo shoot is 8.4 hours and an average of 1.6 photos per tourist.

The trajectory of dual theme tourists is based on relatively complex route patterns at the scale of OTL (Table 3). Mostly the path followed has turning points at POIs with two or more captured photos reflecting the higher attention of dual theme tourists to various attractions. The time of interest of photography starts from 05:00 AM to 06:00 AM with an interval of 1 hour and continues from 08:00 AM to 12:00 AM with further two intervals during 11:00 AM to 11:59 AM and 06:00 PM to 06:59 PM. Most starting points (green flags) are near or above the water streams. The average interval of photography is found to be 23.3 hours and the average number of photoshoots is 3.6 photographs per tourist.

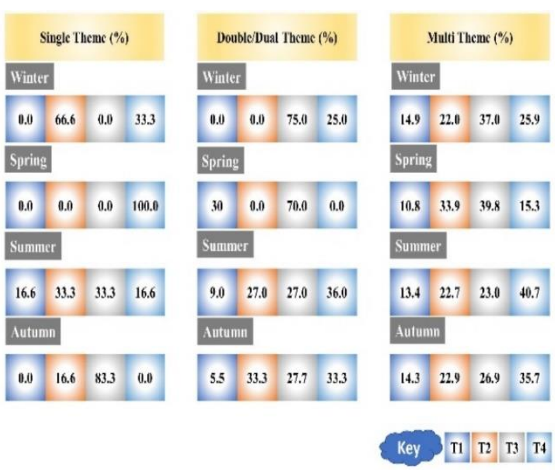

Figure 7. **Spatiotemporal Trajectory of Tourists in the OTL**

IV. DISCUSSION

Examining large collections of images using visual content analysis offers a novel approach to comprehending the inclinations of tourists [36]. Improved comprehension of the spatiotemporal behavior patterns of tourists could aid in managing attractions and improving the overall tourist experience in the Old Town of Lijiang. This analysis employed the convolutional neural network-based model for categorizing visual emotion concepts to retrieve the visual semantics and the spatial and temporal behavior of tourists in the destination images. This model greatly enhances the annotation accuracy in ANPs according to performance evaluation and comparisons with its predecessors [26]. The previously studied visual semantic classification used only one top-ranking result, which does not necessarily characterize the photo subject type accurately. Moreover, we enhance the results of the image deep learning technique innovatively by finding out the mode of nouns in the top five confidence ranking ANPs obtained from recognizing each photo to nominate the visual semantic perceived landscape perception. Another innovative contribution is the classification of tourists according to their diverse photo shoots during destination visits. By analyzing the visual and spatial qualities of tourists' photographs, we outlined their tendencies towards taking pictures and used this information to gain deeper insights into their behaviors.

KDE is a statistical technique utilized in tourism research to estimate the probability density function of a random variable. KDE can be employed in tourism to recognize the spatial concentration of tourist activities and destinations, scrutinize the spatial associations between different tourist destinations and activities, and identify areas that may be at risk of over-tourism or under-tourism. KDE (Figure 4) results showed that 13.46% of the architecture, religion, monuments, and Chinese characters perceived perceptions under historical and cultural landscape (T1) activities. The findings further suggest that this town is a well-planned urban space, with many narrow alleyways, winding streets, and open squares. Therefore, in comparison to historical culture, the people, pets, food, and activities perceptions show more concentration of 24.85% under Urban Spatial Landscape (T4). In contrast to urban space and historical culture, the landscape, meteorology, and environment perceptions are perceived under natural environment landscape (T3), which was found to be 30.64% reflecting the beauty and geographical location of the town surrounded by mountains and forests. Human Life Landscape is the most popular destination representative of OTL, which perceives neighborhoods, spaces, and transportation infrastructure unexpectedly receiving 31.0% of the total attention of tourists under Human Life Landscape (T2). One reason might be the nature of the town to be home to many Naxi people, who continue to preserve their traditional norms of life. It has a rich cultural and artistic heritage with many local crafts, music, and dance traditions with the provision of restaurant services and guesthouses catering to visitors.

The spatiotemporal classification results show the arrangement of tourists based on shooting the landscape perspectives (T1, T2, T3, T4) of the perceived themes. Those who capture a single theme or perspective are called single theme tourists, dual or double theme captures two perspectives and tourists with more than two capturing perspectives of landscape perceptions are nominated as multi-theme tourists. Seasonal-based spatiotemporal classification was conducted using KDE to know about the seasonal behavior of the tourists' inflow and a spatiotemporal trajectory shows the path followed by these tourists.

We have classified the most frequent areas and their behaviors into distinct types of rising, persistent, and declining hotspots [56]. A rising hotspot refers to a location where there is an upward trend in the amount of activity taking place over time. A persistent hotspot is a place where there is no visible trend in the concentration of activities over time. A declining hotspot is a location where there is a downward trend in the amount of activity taking place over time. The KDE (Figure 5) shows the distribution of seasonal activities across different thematic tourists.

Single-theme tourists show a declining temporal hotspot of activities from winter to autumn. Moreover, (Figure 6) explains the highest seasonal proportion based on tourists' photos during winter is T2 (66.66%), spring T4 (100%) and summer earned both T2 and T3 (33.3%), whereas autumn single theme tourists showed highest interest in T3 (83.3%). Unlike single theme, double-theme tourists show a rising temporal hotspot of activities from winter to autumn. The highest seasonal proportion based on tourists' photos during winter is T3 (75%), spring T3 (70%), summer T4 (36%) and autumn earned the highest interest in T2 and T4 (33.3%). Similarly, multi-theme tourists also show rising temporal hotspots of photos, but spring and summer seasons show almost persistent hotspots. The highest proportional seasonal interest of tourists lies in winter season T3 (37%), spring T3 (39.8%), summer T4 (40.7%) and autumn T4 (35.7%). The spatial variation of tourists, perceptions of the visual landscape the perception of single and double-theme visitors' landscape images have "scattered distribution" spatial characteristics, whereas multi-theme landscape perception images have "significant nucleus" spatial characteristics. Multi-theme activities are concentrated along the river streams where double and single theme concentration show specific aims and objectives. The double theme in the summer season presents somehow similar activities to the multi-theme.

According to (Table 2) single-theme, double-theme and multi-theme show all the subordinate tourist types. Moreover, in multi-theme, the highest (23) number of tourists preferred a subordinate type of T1T2T3T4 in all perceptions. According to our study, the proportion of tourists' activities (T1, T2, T3, and T4) of single-theme, double theme and multi-theme tourists are not significantly affected by season but the number of overall activities of thematic tourists significantly vary seasonally from winter to autumn (highest in autumn). Almost similar seasonal results were also found by [57] while studying tourist spatiotemporal behavior in Yuanmingyuan Park. They found consistency across seasons in tourists' behavior. The concentration of the hotspots of tourist highlights is directly proportional to the concentration of services in the destination [58]. Mostly, the density of hotspots in our study is evenly distributed into a belt along the river stream, where shops, markets, and other services are displayed on the map (Figure 5). KDE hotspots were also applied by Peng and colleagues [23] who reported seasonal preferences of tourists at Huangshan using social media data. Our findings are consistent with the seasonal scale observed in the scenic area. Just as tourists at Huangshan focus on different aspects of the landscapes during different seasons, tourists visiting the OTL also have varying preferences throughout the year. For instance, the autumn season at Huangshan, where mountains, rocks and people draw attention. This observation aligns with our findings that autumn is the season when tourists in the OTL capture every aspect of the landscape. Another study was carried out by [36] that found out intercontinental perception differences in Beijing. The authors discovered more interest in North American producing UGC, while European, Asian and Oceania tourists are second, third and fourth number, respectively. KDE-based hotspot maps are also used by [56] for recognizing tourists' and residents' check-in activities using Weibo data. Their conclusion shows most hotspots in the surroundings of shopping malls, nightclubs, and restaurant facilities similar to our results that show the highest density hotspots of multi-theme tourists in services area along the stream channel, where the activities of double theme and single theme are lesser.

Spatiotemporal trajectory characteristics of tourists with different photo-taking behaviors are very crucial. The spatial and temporal structural aspects of geographic data can be considered as an objective method of understanding tourists' socio-economic characteristics up to some extent [21]. According to the tourists' behavioral characteristics, three different flow patterns can be used to summarize the spatial tourist's flow. The spatiotemporal trajectory of single-theme, double-theme, and multi-theme tourists is portrayed (Figure 7). Single themes flow spatial pattern is very simple, where the overall starting points of photo taking are far apart, showing some specific aim of their visits to the OTL with less than two average photoshoots and in shooting interval of 8.4 hours. Most of them lack endpoints inside the OTL, hence following a very simple trajectory shows less enjoyment. Double theme tourists' spatial flow pattern covers comparatively long distances with turns at attention points. During their route pattern, they spend more average time of 23.3 hours and take an average of 3.6 photos per tourist. The trajectory pattern-mining results of multi-theme tourist are very complex and tend to radiate in all directions in OTL. Tourist trajectories intersect multiple times along the river stream showing a more complex route. Multi-theme tourists enjoy an in-depth trip with an average of 66 hours of stay, with an average of more than 47 photos per tourist. The authors in [21] also analyzed tourists' travel paths about geometric features and identified spatiotemporal behavior patterns. They discovered how to use geographic data's spatiotemporal structure to gain insights into the socio-economic characteristics of tourists. Additionally, understanding tourists' socio-economic traits can aid in predicting their spatiotemporal patterns. Single theme tourists' recorded route pattern was from 10:00 AM-10:00 PM, Double theme was from 8:00 AM-11:00 PM and multi-theme was from 7:00 AM to 12:00 AM. The highest number of photo shoots was recorded in multi-theme tourists which were at 11:00 AM and 4:00 PM.

This study presented some limitations that should be addressed. The utilization of the DSB method for analyzing visual semantics of photos, based on Flickr data, aimed to enhance recognition accuracy. However, it is important to note that the generalizability of these results to different photo datasets remains untested. Given the continuous evolution of image recognition technology, it is recommended to integrate more accurate methods in future research to improve recognition outcomes. Furthermore, it is crucial to acknowledge that tourist behavior holds diverse definitions across various fields. Factors such as intrinsic motivation and the representation of tourist destinations significantly influence tourist behavior. The current study primarily focused on analyzing external manifestations observed in tourist photo behavior, which offers a limited scope of understanding. To deepen the analysis of tourist photo behavior, it is necessary to incorporate richer tourist data and consider a broader range of factors. The dataset we used lacked comprehensive attribute information about tourists, including gender, which constrained our ability to thoroughly analyze landscape perceptions and affective preferences among different tourist categories. Moreover, tourists' perceptions of landscapes are influenced by numerous factors beyond visual observations, such as personal experiences and cultural backgrounds. Emotions of visitors are integral to how they experience landscapes, but our dataset only provided partial insights into these emotions. However, our data struggled to represent the intricacies of tourists' interactions with historical, humanistic, natural, and urban spatial landscapes, which significantly impact their experiences. To address these limitations, future research should diversify data sources and consider multicase comparisons to gain a more comprehensive understanding of tourist behavior and preferences. These considerations emphasize the need for further exploration and a comprehensive approach to comprehending and interpreting tourist behavior, particularly about photographs.

V. LIMITATIONS OF THE PROPOSED APPROACH

Despite demonstrating encouraging performance, our proposed methodology has some limitations. DSB was employed for analyze visual semantics of photos from Flickr data aiming at improved recognition accuracy. However, the generalizability of the proposed methodology using more data is still unexplored. We acknowledge that tourists' behavior is a subjective matter that has different meanings in various situations. Intrinsic motivation of the tourists and inherent characteristics of the destinations have significant impact on the tourists' behavior. In this study, we have only focused on the appearances of different scenarios in photos to predict tourists' behavior that limits the scope of understanding. Therefore, we will use more data for in-depth analysis of tourists' behavior. Furthermore, demographic information was not available in the dataset that prohibited a comprehensive analysis of landscape perceptions among various tourists' categories. Similarly, the dataset does not provide other information beyond the visual observations including personal experiences, cultural backgrounds, and social status that may influence landscape perceptions. To address these limitations, we intend to collect diverse and extensive data from multiple sources and consider multicase comparisons for a comprehensive understanding of tourists' behavior and preferences. It is obvious that further exploration is required for better interpretation of tourists' behavior particularly about photographs.

VI. CONCLUSION

This work presents a novel approach to comprehending tourists' preferences from the visual content analysis of pictorial data. We utilized deep learning convolutional neural network image recognition techniques to analyze the OTL as a case study. Specifically, we employed the DSB image recognition model to quantitatively calculate tourists' landscape perceptions and preferences. We analyzed the distribution of tourists based on landscape photoshoots and the factors influencing the variation in thematic tourists' behavior from multiple perspectives. The top 5 visual features computed with DSB have been efficiently exploited using LDA. Further, KDE effectively analyzed the spatial distribution of tourists' behavior at the attraction-level. Based on our analyses, four preferred landscape types among thematic tourists have been identified. It is found that Human Life Landscape earned the most attention, while historical and cultural landscapes earned the least. Additionally, it is observed that tourists' preferences have diversities across the spatiotemporal scale of OTL. For instance, most of the landscape preferences are concentrated in the autumn season of multi-theme tourists along the streams of the river, where single-theme and dual-theme tourists' landscape preferences are distributed at the scale of OTL. By analyzing the visual semantic information extracted from tourists' photos, our findings can provide valuable insights for tourism destination marketing organizations to design targeted strategies for different types of tourists. Furthermore, the visual semantic analysis of photos can reveal the landscape perceptions and preference characteristics of tourists at different spatiotemporal scales, which can contribute to the seasonal management and emergency warnings of scenic spots, the formulation of tourist tour routes during peak hours, and the creation of sustainable tourism products.

List of Abbreviations

ANP Adjective-Noun Pair
AOI Area Of Interest
CNN Convolution Neural Network
DSB DeepSentiBank
EXIF EXchangeable Image File
GPS Global Positioning System
KDE Kernel density estimation
LDA Latent Dirichlet allocation
OTL Old Town of Lijiang
POI Point Of Interest
ReLU Rectified Linear Unit
UGC User-generated content
UNESCO United Nations Educational, Scientific and Cultural Organization

APPENDIX

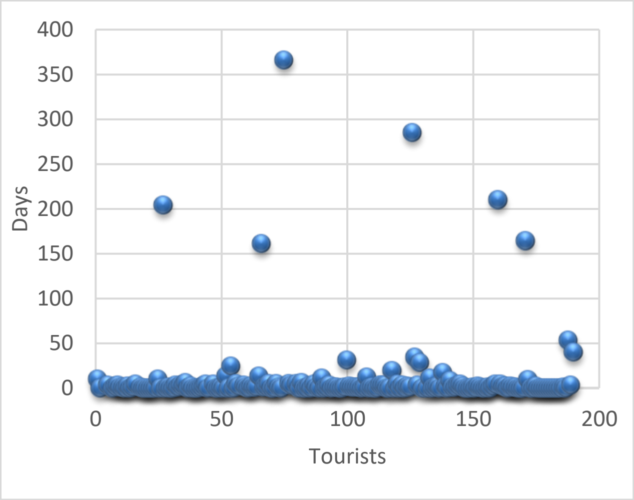

Figure A1. Travel period span of tourists in Lijiang County

ACKNOWLEDGMENT

The authors want to thank Hebei Normal University for the provision of a platform to conduct work on this study.

FUNDING

This research was funded by the National Natural Science Foundation of China grant number 41471127.

CONFLICT OF INTEREST

None declared.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3395469

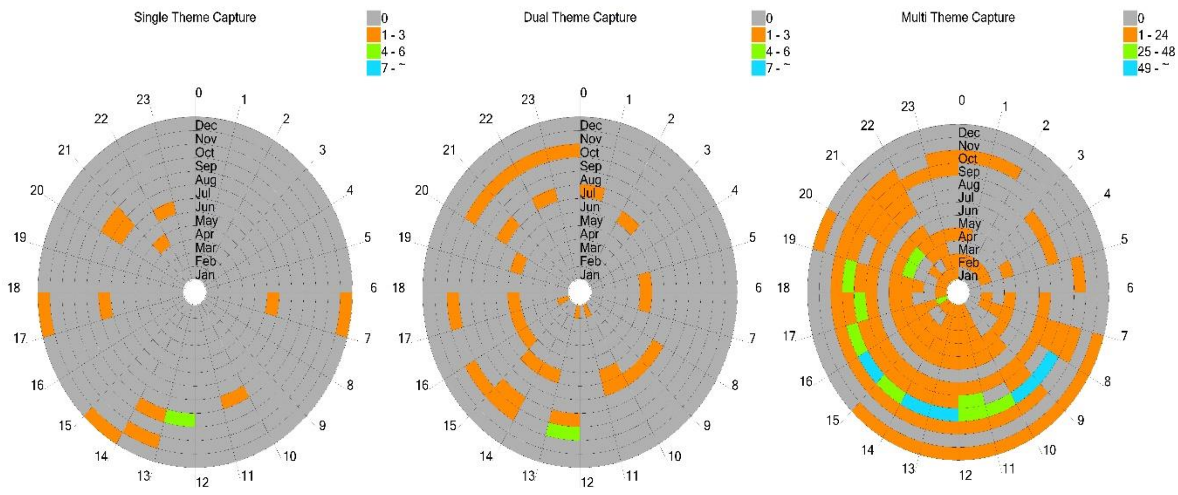

Figure A2. The Clock Diagram shows monthly and hourly activities in OTL.

REFERENCES

[1] C. Jing, M. Dong, M. Du, Y. Zhu, and J. Fu, "Fine-grained spatiotemporal dynamics of inbound tourists based on geotagged photos: A case study in Beijing, China," *IEEE Access*, vol. 8, pp. 28735-28745, 2020.

[2] A. Ghermandi, "Geolocated social media data counts as a proxy for recreational visits in natural areas: A meta-analysis," *J. Environ. Manage.*, vol. 317, pp. 115325, 2022.

[3] R. Huang, "Analyzing national parks visitor activities using geotagged social media photos," *J. Environ. Manage.*, vol. 330, pp. 117191, 2023.

[4] H. Senaratne, A. Mobasheri, A.L. Ali, C. Capineri, and M. Haklay, "A review of volunteered geographic information quality assessment methods," *Int. J. Geogr. Inf. Sci.*, vol. 31, no. 1, pp. 139-167, 2017.

[5] T.T.T. Nguyen and S. Tong, "The impact of user-generated content on intention to select a travel destination," *J. Market. Anal.*, vol. 11, no. 3, pp. 443-457, 2023.

[6] Q. Zhou, M. Sotiriadis, and S. Shen, "Using TikTok in tourism destination choice: A young Chinese tourists' perspective," *Tour. Manag. Perspect.*, vol. 46, pp. 101101, 2023.

[7] N. Deng, L. Zhong, and H. Li, "Destination image perception based on UGC image metadata-Beijing as an example," *Tourism Tribune*, vol. 33, pp. 53-62, 2018.

[8] J. Kim, B. Thapa, S. Jang, and E. Yang, "Seasonal spatial activity patterns of visitors with a mobile exercise application at Seoraksan National Park, South Korea," *Sustainability*, vol. 10, no. 7, pp. 2263, 2018.

[9] M. Sinclair, M. Mayer, M. Woltering, and A. Ghermandi, "Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks," *J. Environ. Manage.*, vol. 263, pp. 110418, 2020.

[10] Y. Jiansheng and L. Yanling, "Research on Spatial-Temporal Behavior Characteristics of Tourists Based on Digital Footprints: A Case Study of Guilin, China," *J. Phys-Conference Series*, vol. 1883, no. 1, p. 012058, IOP Publishing, 2021.

[11] J. Mou, "Extracting Network Patterns of Tourist Flows in an Urban Agglomeration Through Digital Footprints: The Case of Greater Bay Area," *IEEE Access*, vol. 10, pp. 16644-16654, 2022.

[12] Z. Shao, N.S. Sumari, A. Portnov, F. Ujoh, W. Musakwa, and P.J. Mandela, "Urban sprawl and its impact on sustainable urban development: a combination of remote sensing and social media data," *Geo-Spat. Inf. Sci.*, vol. 24, no. 2, pp. 241-255, 2021.

[13] H.Q. Vu, G. Li, R. Law, and Y. Zhang, "Tourist activity analysis by leveraging mobile social media data," *J. Travel Res.*, vol. 57, no. 7, pp. 883-898, 2018.

[14] W. Liu, B. Wang, Y. Yang, N. Mou, Y. Zheng, L. Zhang, and T. Yang, "Cluster analysis of microscopic spatiotemporal patterns of tourists' movement behaviors in mountainous scenic areas using open GPS-trajectory data," *Tourism Manage.*, vol. 93, pp. 104614, 2022.

[15] Y. Zheng, N. Mou, L. Zhang, T. Makkonen, and T. Yang, "Chinese tourists in Nordic countries: An analysis of spatiotemporal behavior using geo-located travel blog data," *Comput. Environ. Urban Syst.*, vol. 85, pp. 101561, 2021.

[16] S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, R. Shibasaki, "Understanding tourist

behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan," *Pervasive Mob. Comput.*, vol. 18, pp. 18-39, 2015.

[17] Y. Liu, J. Shi, and M. Jian, "Understanding visitors' responses to intelligent transportation system in a tourist city with a mixed ranked logit model," *J. Adv. Transp.*, vol. 2017, 2017.

[18] Y. Li, L. Yang, H. Shen, and Z. Wu, "Modeling intra-destination travel behavior of tourists through spatiotemporal analysis," *J. Destin. Mark. Manag.*, vol. 11, pp. 260-269, 2019.

[19] H. Han, S. Kim, and F.E. Otoo, "Spatial movement patterns among intra-destinations using social network analysis," *Asia Pac. J. Tour. Res.*, vol. 23, no. 8, pp. 806-822, 2018.

[20] X. Zhao, X. Lu, Y. Liu, J. Lin, and J. An, "Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an, China," *Tourism Manage.*, vol. 69, pp. 368-383, 2018.

[21] X. Huang, M. Li, J. Zhang, L. Zhang, H. Zhang, and S. Yan, "Tourists' spatial-temporal behavior patterns in theme parks: A case study of Ocean Park Hong Kong," *J. Destin. Mark. Manag.*, vol. 15, pp. 100411, 2020.

[22] A. Birenboim, S. Anton-Clavé, A.P. Russo, and N. Shoval, "Temporal activity patterns of theme park visitors," *Tour. Geogr.*, vol. 15, no. 4, pp. 601-619, 2013.

[23] J. Yan, J. Yue, J. Zhang, and P. Qin, "Research on Spatiotemporal Characteristics of Tourists' Landscape Perception and Emotional Experience by Using Photo Data Mining," *Int. J. Environ. Res. Public Health*, vol. 20, no. 5, pp. 3843, 2023.

[24] Z. Li, Y. Fan, W. Liu, and F. Wang, "Image sentiment prediction based on textual descriptions with adjective noun pairs," *Multimed. Tools Appl.*, vol. 77, pp. 1115-1132, 2018.

[25] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[26] T. Chen, D. Borth, T. Darrell, and S-F. Chang, "Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks," *arXiv preprint arXiv :1410.8586*, 2014.

[27] D. Xu, L. Cong, and G. Wall, "Tourists' spatiotemporal behaviour and concerns in park tourism: Giant Panda National Park, Sichuan, China," *Asia Pac. J. Tour. Res.*, vol. 24, no. 9, pp. 924-943, 2019.

[28] N. Shoval and M. Isaacson, "Sequence alignment as a method for human activity analysis in space and time," *Ann. Assoc. Am. Geogr.*, vol. 97, no. 2, pp. 282-297, 2007.

[29] T. Zhang, Z. Lian, and Y. Xu, "Combining GPS and space syntax analysis to improve understanding of visitor temporal–spatial behaviour: a case study of the Lion Grove in China," *Landsc. Res.*, vol. 45, no. 4, pp. 534-546, 2020.

[30] W.C. Hunter, "The social construction of tourism online destination image: A comparative semiotic analysis of the visual representation of Seoul," *Tourism Manage.*, vol. 54, pp. 221-229, 2016.

[31] X. Xiao, C. Fang, H. Lin, and J. Chen, "A framework for quantitative analysis and differentiated marketing of tourism destination image based on visual content of photos," *Tourism Manage.*, vol. 93, p. 104585, 2022.

[32] J. Qin, C. Song, M. Tang, Y. Zhang, and J. Wang, "Exploring the spatial characteristics of inbound tourist flows in China using geotagged photos," *Sustainability*, vol. 11, no. 20, p. 5822, 2019.

[33] Y. Liu and W. Liao, "Spatial characteristics of the tourism flows in China: A study based on the Baidu index," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 6, p. 378, 2021.

[34] C.-L. Kuo, T.-C. Chan, I.-C. Fan, and A. Zipf, "Efficient method for POI/ROI discovery using Flickr geotagged photos," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 3, p. 121, 2018.

[35] N. Mou, R. Yuan, T. Yang, H. Zhang, J.J. Tang, and T. Makkonen, "Exploring spatiotemporal changes of city inbound tourism flow: The case of Shanghai, China," *Tourism Manage.*, vol. 76, p. 103955, 2020.

[36] K. Zhang, Y. Chen, and C.J.T.M. Li, "Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing," *Tourism Manage.*, vol. 75, pp. 595-608, 2019.

[37] N. Deng, J. Liu, Y. Niu, and W.X. Ji, "Different perceptions of Beijing's destination images from tourists: An analysis of Flickr photos based on deep learning method," *Resour. Sci.*, vol. 41, no. 3, pp. 416-429, 2019.

[38] X. Huang, Y. Wang, J. Li, X. Chang, Y. Cao, J. Xie, and J. Gong, "High-resolution urban land-cover mapping and landscape analysis of the 42 major cities in China using ZY-3 satellite images," *Sci. Bull.*, vol. 65, no. 12, pp. 1039-1048, 2020.

[39] M. Bubalo, B.T. Van Zanten, and P.H. Verburg, "Crowdsourcing geo-information on landscape perceptions and preferences: A review," *Landsc. Urban Plan.*, vol. 184, pp. 101-111, 2019.

[40] A.J.L. Dunkel, "Visualizing the perceived environment using crowdsourced photo geodata," *Landsc. Urban Plan.*, vol. 142, pp. 173-186, 2015.

[41] R.W. Figueroa-Alfaro, Z.J. Tang, "Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr," *J. Environ. Plan. Manag.*, vol. 60, no. 2, pp. 266-281, 2017.

[42] Z. Guo and L. Sun, "The planning, development and management of tourism: The case of Dangjia, an ancient

village in China," *Tourism Manage.*, vol. 56, pp. 52-62, 2016.

[43] W.J. Zheng, "Research on cruise tourists' expectations and perceptions of Li River landscape," *China Popul. Resour. Environ.*, vol. 23, pp. 143-148, 2013.

[44] G. Parasher, M. Wong, M. Rawat, "Evolving role of artificial intelligence in gastrointestinal endoscopy," *World J. Gastroenterol*., vol. 26, no. 46, pp. 7287, 2020.

[45] G. del Castillo Torres, M.F. Roig-Maimó, M. Mascaró-Oliver, E. Amengual-Alcover, and R. Mas-Sansó, "Understanding How CNNs Recognize Facial Expressions: A Case Study with LIME and CEM," *Sensors*, vol. 23, no. 1, p. 131, 2022.

[46] S. Stepchenkova and F. Zhan, "Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography," Tourism Manage., vol. 36, pp. 590-601, 2013.

[47] J. Li, T. Dai, S. Yin, Y. Zhao, D.I. Kaya, and L. Yang, "Promoting conservation or change? The UNESCO label of world heritage (re) shaping urban morphology in the Old Town of Lijiang, China," *Front. Archit. Res.*, vol. 11, no. 6, pp. 1121-1133, 2022.

[48] B. Thomee, D.A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.J. Li, "YFCC100M: The new data in multimedia research," *Commun. ACM*, vol. 59, no. 2, pp. 64-73, 2016.

[49] P.J. Bell, "Content analysis of visual images," *Handbook of Visual Analysis*, vol. 13, 2001.

[50] A.H.N. Mak, "Online destination image: Comparing national tourism organisation's and tourists' perspectives," *Tourism Manage.*, vol. 60, pp. 280-297, 2017.

[51] D. Newman, P. Smyth, M. Welling, and A. Asuncion "Distributed inference for latent dirichlet allocation," *Adv. Neural Inf. Process. Syst.*, vol. 20, 2007.

[52] F. Yan, N. Xu, and Y. Qi, "Parallel inference for latent dirichlet allocation on graphics processing units," *Adv. Neural Inf. Process. Syst.*, vol. 22, 2009.

[53] A. Okabe, T. Satoh, and K. Sugihara, "A kernel density estimation method for networks, its computational method and a GIS - based tool," *Int. J. Geogr. Inf. Sci.*, vol. 23, no. 1, pp. 7-32, 2009.

[54] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, Jan., pp. 993-1022, 2003.

[55] S.Y. Han, M.-H. Tsou, and K.C. Clarke, "Do global cities enable global views? Using Twitter to quantify the level of geographical awareness of US cities," *PloS One*, vol. 10, no. 7, art. e0132464, 2015.

[56] X. Su, B. Spierings, M. Dijst, and Z. Tong, "Analyzing trends in the spatiotemporal behavior patterns of mainland Chinese tourists and residents in Hong Kong based on Weibo data," *Curr. Issues Tour.*, vol. 23, no. 12, pp. 1542-1558, 2020.

[57] Q. Yao, Y. Shi, H. Li, J. Wen, J. Xi, and Q. Wang, "Understanding the tourists' Spatiotemporal behavior using open GPS trajectory data: A case study of yuanmingyuan park (Beijing, China)," *Sustainability*, vol. 13, no. 1, p. 94, 2020.

[58] G. Lau and B. McKercher, "Understanding tourist movement patterns in a destination: A GIS approach," *Tour. Hosp. Res.*, vol. 7, no. 1, pp. 39-49, 2006.

Muhammad Iqbal has completed his master's degree in Geography from University of Peshawar. Currently, he is perusing his PhD degree in Geographic Information System from Hebei Normal University, People's Republic of China. His research interests include GIS, Tourism and Planning Management, Data Mining, and Data visualization.