A study of landscape preference formation mechanism based on convolutional neural network and Grad-CAM

Jingxuan Hu[1], Zhijie Yang[2] and Yujie Ren[*3]

[1]Undergraduate student, Department of Urban and Rural Planning, Nanjing Forestry University, China
[2]Undergraduate student, Department of Landscape Architecture, Nanjing Forestry University, China
[3]Lecture, Department of Urban and Rural Planning, Nanjing Forestry University, China

Abstract

Landscape preference arises from the process of human interpretation and judgment of landscape, and an accurate understanding of urban residents' landscape preferences and their causes can help decision makers formulate planning policies and improve design satisfaction. However, traditional studies on landscape preference are mainly based on the premise of direct/indirect contact between observers and the landscape, which not only consumes a lot of costs, but also cannot quantitatively explain the formation path of landscape preferences. Therefore, this study proposes a method to characterize landscape preferences based on online word-of-mouth data and machine learning, and to analyze the formation mechanism of landscape preferences. By extracting the landscape evaluations contained in online word-of-mouth data, convolutional neural network models for landscape preference prediction are trained to measure the degree of landscape preference quantitatively. The corresponding landscape preference templates are also generated. Based on the results of landscape preference analysis, this study further utilizes the convolutional feature map and Grad-CAM to explain the basis of the CNN models to predict landscape preferences and clarify their formation mechanism.

The output of this study is expected to improve the theories related to landscape preference on the basis of significantly reducing the cost of landscape preference acquisition.

Keywords: landscape preference; convolutional neural network; machine learning; Grad-CAM

1 Introduction

In the context of public participation in high-quality urban renewal, how to ensure that design solutions meet the landscape aesthetic and behavioral preferences of urban residents with different attributes has become a critical issue to be addressed in the field of urban design. It is believed that landscape preference arises from the interaction between people and the environment, and is a subjective evaluation process in which viewers interpret and judge the landscape through attention selection, personal social background, culture, and psychological state, thus forming a cognitive and preferential degree. Individuals with different social life backgrounds, physiological conditions and states may differ greatly in their preferences for the same landscape. In a diverse urban population composition pattern, accurately obtaining the landscape preferences of urban residents with different attributes can help decision makers to formulate planning policies and improve urban design satisfaction.

In order to accurately obtain the landscape preferences of urban residents, researchers from the fields of urban planning, landscape design, geography and ecology have carried out a lot of analysis and measurement work on them by adopting various methods. The traditional landscape preference characterization methods include (1) active choice based analysis methods such as visitor employed photography (VEP), (2) passive choice based analysis methods such as scenic beauty evaluation (SBE) and questionnaire, (3) subconscious choice based analysis methods such as eye-movement test, (4) direct observation based behavior recording method, behavior observation, behavior map, and behavior image (4) behavior intention analysis method based on direct observation such as behavior recording method, behavior observation, behavior map and behavior image, (5) track behavior tracking method based on location identification technology such as positioning system and tracking data analysis[1]. Based on the questionnaire feedback, photo commentary and usage behavior records of urban residents with different attributes, the landscape preference feature analysis methods based on photography, questionnaires, tracking

***Contact Author:** Yujie Ren, Lecture, Nanjing Forestry University
No.159 Longpan Road, Nanjing, 210037, Jiangsu China
Tel.: 086+18651674270
E-mail: renyujie@njfu.edu.cn

research and other traditional methods might consume a lot of labor, time and material costs while accurately analyzing the landscape preferences of urban residents with different attributes. For example, the tourist hired photography method, the beauty degree evaluation method and the questionnaire method do not require much professional equipment and are easy to operate, but the implementation process requires days or even weeks to arrange professional staff to organize relevant experiments; the eye-movement test method, global positioning system, tracking data analysis and other methods can study human preferences more objectively, but require more professional equipment and a research management platform that can obtain a large amount of tracking data; the behavioral observation method The behavioral observation method, behavioral map method and other methods can use simple tools to observe and study people's objective preferences, but it is more difficult to observe continuously for a long time[2].

Based on this, the main task of this research is to design a new method to quantitatively measure and analyze the evaluation and preference characteristics of urban residents with different attributes for unfamiliar urban landscapes and to elucidate their formation mechanisms, using the expertise of machine learning in the field of prediction without questionnaires, field research and data tracking.

2 Literature Review

2.1 Overview of studies related to landscape preference characterization methods

The mainstream landscape preference characterization methods currently rely on photography, questionnaires and tracking research. The main methods used include active choice based evaluation methods such as tourist employed photography (VEP), passive choice based evaluation methods such as scenic beauty evaluation (SBE) and questionnaire and track behavior tracking methods based on location identification technology such as tracking data analysis. These methods mainly analyze landscape preference characteristics by directly or indirectly contacting landscape observers to obtain his/her/their evaluation information about a specific urban landscape. In the process of quantifying landscape evaluation information, the three methods of environmental preference matrix, environmental evaluation scale and photo scoring proposed by the Kaplan couple were the first research tools that attempted to quantify landscape preference evaluation and were also widely used in later studies. In recent years, web surveys in the context of the emerging Internet revolution have been used to obtain landscape preference data, and their reliability has been demonstrated[1]. With cell phone photography, video capabilities, and mobile applications making public environmental impression ratings proliferate on the Internet, there has been a trend to use public self-reporting, social media, and crowdsourced geographic information to obtain perception and preference data. These studies have successfully addressed how to collect and model people's affective responses to the environment into landscape preference studies by actively asking users to rate the level of comfort, safety, diversity, attractiveness, and relaxation experiences in the environment on social platforms or mobile apps to collect users' affective responses based on geographic information.

2.2 Overview of studies related to factors influencing landscape preference

In terms of factors influencing landscape preference, the study points out that since landscape preference is a product of perception, the outcome of human interaction with landscape should be influenced by the subject of landscape aesthetic experience (individual or group attributes of the observer) and the landscape object (landscape type, visual space attributes, landscape quality, etc.). Aesthetic subject characteristics mainly include individual or group attributes of the observer. Such studies start from the attributes of the subjects of landscape experience and explore the influence of the characteristics of respondents' occupation, gender, age, race, familiarity, personality, etc., as well as the group characteristics of local farmers, residents, tourists, students, etc., on landscape preferences[2]. The above studies yielded some representative conclusions: occupational background, education level and gender have a greater impact on landscape preferences compared to other demographic characteristics. Landscape object characteristics mainly include landscape quality, landscape type, and visual-spatial attributes. The study showed that the higher the landscape quality, the stronger the consistency of preference. Landscape type also has a significant impact on the consistency of landscape preference judgments. Compared with artificial landscapes, natural landscapes are commonly preferred.

2.3 Future trends in the field of landscape preference research

In the direction of landscape preference feature analysis, how to effectively reduce the cost of landscape preference research and whether it is possible to predict the preference of unknown and unfamiliar landscapes without generating contact with landscape evaluation subjects are potential development trends in the field of landscape preference characterization in the future.

In the direction of influencing factors of landscape preference, quantitatively and deeply excavating the influence of the subject, composition, and spatial characteristics of visual elements on landscape preference is expected to provide a new perspective for elucidating the relevant research theories on the formation mechanism of landscape preference, and is gradually developing into a future hot spot in the field of landscape preference research.

3 Landscape preference analysis based on machine learning

3.1 Experimental design

(1) Subjective landscape evaluation information extraction based on online word-of-mouth data

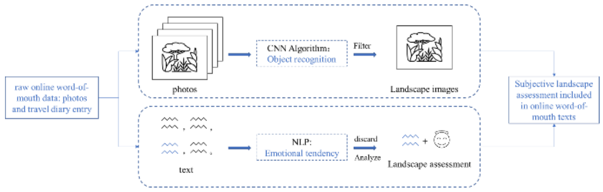


Fig.1 Process of landscape evaluation information extraction

For each set of raw data collected from online word-of-mouth (eWOM) platform for cityscape evaluation in text and image formats, ① use object recognition algorithms in the field of computer vision to screen the parts of online eWOM platform's travelogues texts and tourists' photos that take cityscape as the main subject of the evaluation; ② use the opinion extraction and sentiment tendency analysis algorithms in the field of natural language processing to compute the sentiment information contained in the screened landscape review texts (positive/negative); and ③ extract the review publisher's subjective evaluation of the cityscape corresponding to the text (like/dislike).

(2) Construction of predictive models for landscape evaluation based on machine learning

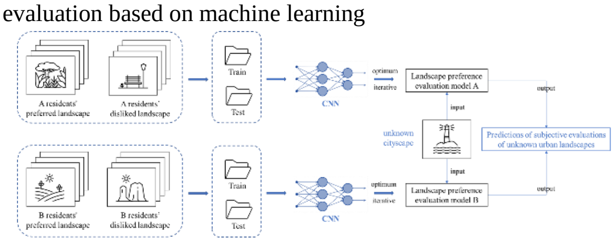


Fig.2 Process of predictive model construction

The extracted subjective landscape evaluation results of urban residents with different attributes are grouped according to their personal attributes and evaluation results, and further ①build a basic database for machine learning that reflects the landscape preferences of urban residents with different attributes; ② divide the training set and test set of the landscape preference prediction machine learning model, and reasonably expand the capacity of the dataset; ③ build a landscape preference prediction machine learning model (convolutional neural network model), and optimize the training parameters and iteration structure of the model to improve the accuracy of the model prediction results; ④Input the unknown city landscape photos into the prediction model, record the output results of the prediction model, and get the subjective evaluation (like/dislike) of the city residents on the unknown city landscape.

(3) Characterization of landscape preferences based on double cross-structural modeling

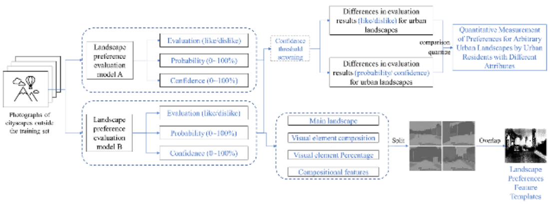


Fig.3 Process of landscape preferences standardization

Using the constructed landscape evaluation prediction model for urban residents with different attributes, ①input a large number of photos of urban landscapes other than the model training set into the landscape evaluation prediction model for urban residents with different attributes, and record the prediction results of the model's emotional tendency (like/dislike) as well as the corresponding output parameters: "likelihood (%)" and "confidence level (%)". ②Filter the subjective evaluation prediction results (like/dislike) of residents with different attributes on urban landscape and the corresponding "likelihood" parameter when the "confidence" parameter reaches the threshold or above. Filter the subjective evaluation prediction results (like/dislike) of residents with different attributes to the cityscape and the corresponding "possibility" parameter values under the scenarios where the "confidence level" parameter reaches the threshold or above, and record them as the subjective landscape evaluation prediction results. Summarize the results, and complete the quantitative measurement of the preference of urban residents of different attributes for any city landscape (in percentage form, %). The feature extraction and convolutional filtering processes of convolutional neural network are used in a cyclic manner to analyze the landscape types, composition/percentage of various visual elements, and the compositional characteristics of each cityscape photo in datasets.

(4) Analyzing the mechanism of landscape preference formation from the perspective of the composition of landscape visual elements

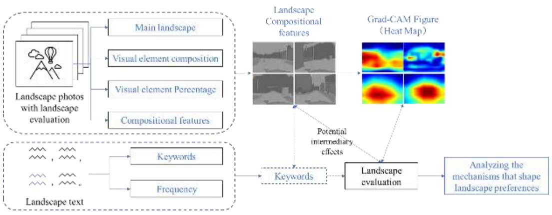


Fig.4 Process of preference formation mechanism analysis

Based on the results of landscape preference characterization, ①with the help of statistical models, analyze the relationship between the main landscape types, the composition/percentage of each type of visual elements, and the compositional features of urban landscape and the quantitative measurement results of the preference of urban residents with different attributes, and analyze the quantitative relationship between the composition of the visual elements of the landscape and the preference of urban landscape; ② through the extraction of the feature maps produced by the convolutional layer during the process of predicting the landscape preference generated by the convolutional neural network in the process of predicting landscape preferences, the key pixels affecting the preference prediction made by the neural network model will be explored. By analyzing the characteristics of the visual elements of the landscape in the region of the key pixels, the mechanism of the formation of landscape preferences will be elucidated.

Up to this point, this research has realized the characterization results (quantitative measurement percentage results + preference templates) of the preferences of city residents of different attributes for any unknown city landscapes without questionnaires, interviews and tracking research, and has made a preliminary explanation of the formation mechanism of landscape preferences from the perspective of the composition of landscape visual elements, just by analyzing the city landscape review data generated by different attributes of city residents in the online word-of-mouth platforms.

3.2 Data acquisition and preprocessing

(1) As for the data collection content, the raw data to be used in this research are the word-of-mouth (eWOM) reviews of Nanjing (China)'s representative urban landscapes by permanent residents and tourists, which are categorized into: (1) online eWOM data in picture format (i.e., tourists' photographs); (2) textual review information (i.e., travelogues' texts); and (3) personal residence attributes of the eWOM publishers.

(2) Regarding the data collection platform, the platform for collecting raw data is TripAdvisor. It is the world's leading travel website with branches in 49 markets around the world, with a monthly average of 415 million unique visitors. 2019 is the same as that of the domestic Ctrip Group. In 2019, the company announced a strategic partnership with Ctrip Group, which formally accessed a large amount of Chinese online word-of-mouth (eWOM) review data generated on Ctrip.com.

(3) With regard to the data collection period, the we had completed the collection of word-of-mouth review information generated by TripAdvisor platform in representative urban landscapes of Nanjing before 2019 (see in references 3) in the preliminary work of this research. The collected data is released between January 2012 and December 2019.

(4) With regard to the data collection area, the applicant has previously studied the spatial distribution characteristics of social network user check-in data, and found that from the perspective of landscape representativeness, there exist a total of 41 representative sites in Nanjing (see references 3 in the references section of the study for details). Therefore, the data collection area of this research is defined as the area corresponding to these 41 representative cityscapes (Fig.5). In the data collection process, the names of these representative cityscapes will be sequentially entered into the search engine of the TripAdvisor platform for the data collection work.

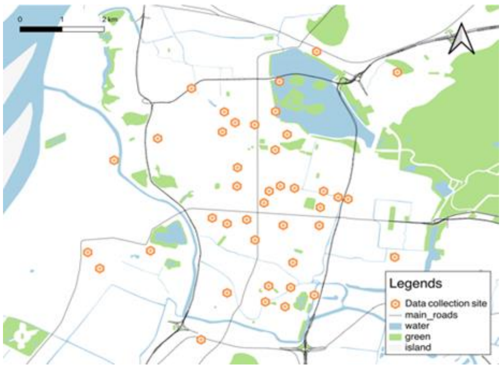


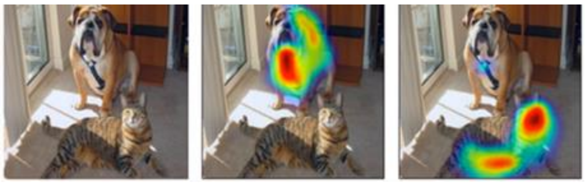
Fig.5 Data collection sites

(5) As for the data collection method, this research will use the WebScraper tool and the Java language to automatically parse the corresponding CSS paths of the web page elements on the TripAdvisor platform, combine the nesting relationship and data structure between the elements, and recursively obtain the online word-of-mouth data in the web page in the form of pictures, the textual comment information paired with them, and the personal attribute of the publisher of the word-of-mouth in page-by-page fashion and store the data in the format of jpg and csv.

After collecting the data, we further introduce CNN and NLP technology to filter the raw data. From 4552 raw photos, 1290 images with urban landscape as the main object and corresponding text comments containing clear subjective comments on the landscape were eventually extracted for constructing the prediction models.

3.3 Core technology and algorithms: Grad-CAM

Deep network visualization method based on gradient localization (Grad-CAM): the algorithm analyzes the convolutional neural network in the last global average pooling layer (Pooling layer) before the category activation mapping picture is generated into the cumulative weighted activation, calculates the weight of the last convolutional layer of each feature map on the category of the picture, and then seeks the weighted sum of each feature map, and finally puts the weighted sum of the feature map on the last convolutional layer. Finally, the weighted sum of feature maps is mapped to the original picture to explain the classification basis of the deep neural network model in the form of a heat map, and the category judgment is made by the pixels of the picture (Fig. 6). This algorithm is the key to analyze the formation mechanism of landscape preference from the perspective of visual element features in this study, and the algorithm relies on the Grad-CAM model implementation of the Pytorch library.



(a) Raw picture (b)Visualization of the "dog" recognition process by Grad-CAM (c)Visualization of the "cat" recognition process by Grad-CAM

Fig.6 Schematic diagram of grad-cam principle

4 Experimental results

4.1 Construction and performance of the models

The 1290 filtered photos with subjective sentiment labels were classified into 12 groups according to the publisher's attributes and label types (Tab.1). Then, we divided these 1290 photos into training data and test data of landscape preference prediction model according to 80% to 20%. Specifically, if 80% of the total number of each group of photos is not an integer, we round it down and record the results as training data, and the rest as testing data. As a result, we input 1016 of the photos as training data into 6 CNN architectures to generate 6 prediction models which can represent the cognitive preference for cultural and natural landscape of local residents, domestic tourists and foreign tourists, respectively. The remaining 274 photos will be used as testing data for model accuracy test.

Tab.1 Model training and testing sets

Group	Attribute	Type	Label	Training (80%)	Testing (20%)	Sum
1	Local Residents	Natural Landscape	Positive	118	30	148
2			Negative	28	10	38
3		Cultural Landscape	Positive	140	36	176
4			Negative	36	11	47
5	Domestic Tourists	Natural Landscape	Positive	160	40	200
6			Negative	40	11	51
7		Cultural Landscape	Positive	196	51	247
8			Negative	48	14	62
9	Foreign Tourists	Natural Landscape	Positive	118	30	148
10			Negative	28	11	39
11		Cultural Landscape	Positive	84	23	107
12			Negative	20	7	27
Total amount				1016	274	1290

It is worth mentioning that due to the lack of data volume corresponding to each label in the input data set, we used the method of data augmentation (add noise, blur, expose, flip and rotate) to expand the size of the training data set (Fig.7). After each original image is optimized by five data enhancement tools (cross-use and superpose-use) in Apple Core ML architecture, the overall volume of the database could meet the training data requirements of the image classifier model based on CNN network.

Finally, we trained 6 landscape preference prediction models based on CNN network. The accuracy of the 6 urban scene preference prediction models built in this study ranged from 98.0% to 100% and 84.4% to 96.8% at 25th iteration in the training data set and the testing data set (Fig.8), respectively. Specifically, the

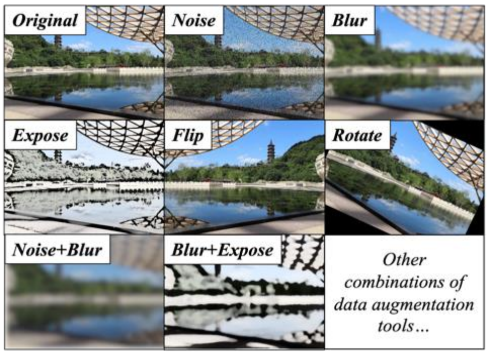


Fig.7 Data augmentation

model constructed for predicting domestic tourist's preference on natural landscapes showed the highest accuracy in both training and validation, while that on cultural landscapes is the lowest.

Tab.2 Model training and testing results

Label	L-P-p.	L-N-p.	L-Cons.	D-P-p.	D-N-p.	D-Cons.	F-P-p.	F-N-p.	F-Cons.
L-NL-P (148)	——			120	11	81.1%	123	4	83.1%
L-NL-N (38)				2	28	73.7%	5	22	57.9%
L-CL-P (176)				129	8	73.3%	141	6	80.1%
L-CL-N (47)				3	31	66.0%	11	28	59.6%
D-NL-P (200)	150	14	75.0%	——			139	13	69.5%
D-NL-N (51)	2	37	72.6%				4	31	60.8%
D-CL-P (247)	193	14	78.1%				187	13	75.7%
D-CL-N (62)	6	50	80.7%				5	46	74.2%
F-NL-P (148)	116	7	78.4%	104	24	70.3%	——		
F-NL-N (39)	8	27	69.2%	13	16	41.0%			
F-CL-P (107)	79	14	73.8%	74	13	69.2%			
F-CL-N (27)	4	14	51.9%	6	19	70.4%			

**Note: L, D, and F denote local residents, domestic tourists and foreign tourists, respectively; NL and CL denote natural landscape and cultural landscape, respectively; P and N denote positive and negative subjective sentiment, respectively; Cons. denotes the consistency between the sentiment label and prediction results; p. denotes that the prediction results with confidence value over 95%.*

In general, since the accuracy of landscape preference prediction models reached 99.55% and 91.07% in the process of training and validation averagely, the models can accurately predict the subjective preferences of three groups of people for different types of urban landscapes (Tab.2).

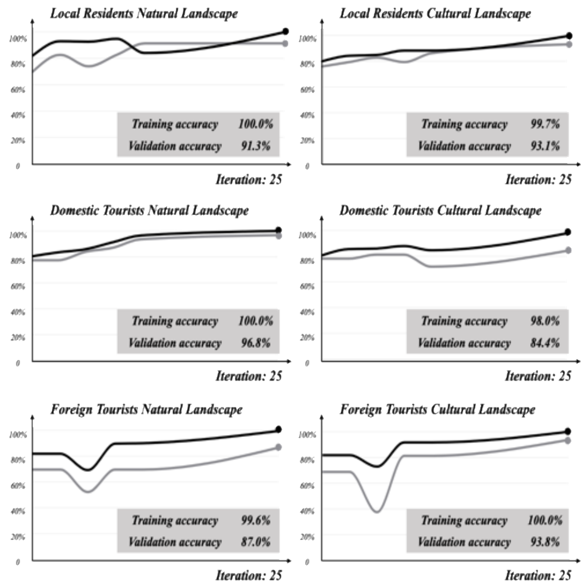


Fig.8 Model training process

Based on the correlation between the percentage (%) of the type (Label) of each visual element and the value of the Probability parameter when the Confidence Level reaches a threshold value of 95% or more, a CNN-based Grad-CAM model (based on the python language) is constructed to derive the mechanism of landscape preference formation for the permanent residents and tourists of Nanjing from the point of view of the composition of the visual elements of the landscape.

The results showed that (Fig.9, the pictures in rows 1-4, 5-8, and 9-12 of the figures show the results of the Grad-CAM model analysis of landscape preferences of foreign, local, and foreign tourists respectively) the main landscape contents, elements and composition characteristics of urban scenes are closely related to the subjective preferences of three groups of people on urban landscapes. Specifically, the unique style of statues, rockeries can make the natural landscape favored by domestic tourists, while foreign tourists will pay more attention to the open grasslands when evaluating natural landscapes. In the cultural landscape, historic doors, murals and decorations with prominent colors (i.e., golden) are the key factors to attract domestic tourists and foreign tourists, respectively.

5 Conclusions

In summary, with the help of online word-of-mouth platform review data and machine learning algorithms, this study analyzes the landscape preference characteristics of urban residents, tries to elucidate the core influencing factors of landscape preference and its formation mechanism from perspective of the composition of landscape visual elements, and is expected to improve the theories in the field of landscape preference research.

It also presents the innovative application scenarios of existing mature machine algorithms in the practice of urban renewal projects, and explores the change and development potential of algorithms such as natural language processing and computer vision for urban planning and landscape design, which is expected to provide a scientific paradigm for the research of the application scenarios of other machine algorithms in the field of urban design, both in terms of methodology and process.

References

[1] Jorgensen A . Beyond the view: Future directions in landscape aesthetics research[J]. Landscape and Urban Planning, 2011,100(04):353-355.

[2] Dunkel, Alexander. Visualizing the perceived environment using crowdsourced photo geodata[J]. Landscape & Urban Planning, 2015, 2(142):173-186.

[3] Yujie Ren, Shichen Zhao. A Study on the Analysis Method of Cognition Preferences for Urban Landscapes Based on SNS Data and Machine Learning[J]. AIJ Journal of Technology and Design (AIJT), 2022,28:929-934. *(Preceding research for this paper)*

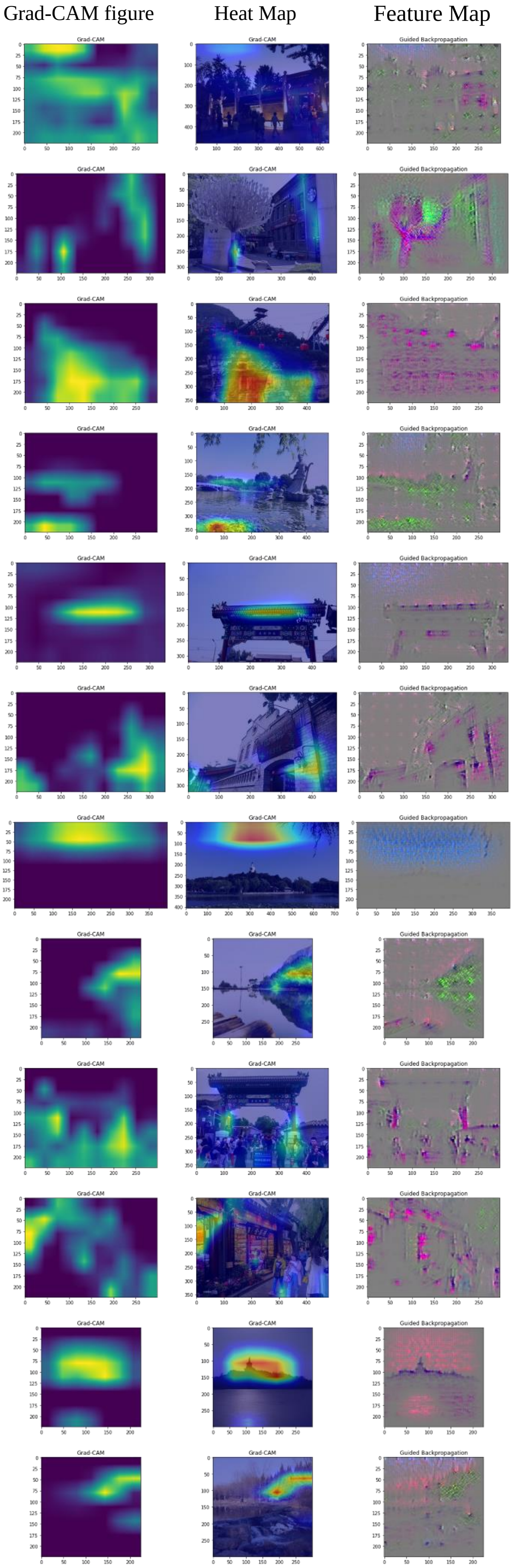


Fig.9 Grad-CAM convolutional layer visualization and analysis map