Hindawi
Mobile Information Systems
Volume 2022, Article ID 5934071, 9 pages
https://doi.org/10.1155/2022/5934071

Research Article

Behavior Recognition and Exhibition Tour Scene Classification Using In-Depth Learning

Xinlu Ren

College of Finance and Economics, Guangzhou Huashang College, Guangzhou, Guangdong 511300, China

Correspondence should be addressed to Xinlu Ren; 20210029@hstc.edu.cn

Received 18 April 2022; Revised 13 May 2022; Accepted 25 May 2022; Published 8 June 2022

Academic Editor: Muhammad Babar

The rapid development of exhibition tourism has led to a sharp increase in the amount of data in the tourism and exhibition industry. Through in-depth mining and application of exhibition tourism data, it can intuitively show its potential relevance and produce much valuable knowledge. The huge value of exhibition tourism data can better meet social needs. Through the analysis of exhibition tourism data, it has certain use-value and significance for the development of the industry. Scene classification in the field of computer vision is a research hotspot. However, there are far few research-related algorithms on mice tourism scene classification. Therefore, based on in-depth learning, this paper studies behavior recognition, and mice tourism scene classification, applies computer vision technology to mice tourism scene classification, collects many visual data, and speeds up the rapid development of the field of vision. In this paper, the scene classification algorithm based on a self-attention generation countermeasure network is constructed to deal with the problem of convention and exhibition tourism. The test results show that the accuracy of the classification results of this algorithm is as high as 99.12%, which is the highest compared with the classification results of other models, which fully proves that this algorithm can accurately classify conventional and exhibition tourism scenes.

1. Introduction

In recent years, with the rapid development of artificial intelligence technology, more and more researchers use computers to simulate biological nerves. Artificial intelligence technology on human behavior recognition is also widely used in smart homes, video monitoring, and other fields [1]. Today, the amount of data in MICE scenarios is increasing sharply, which contains a lot of useful features to be extracted. This paper uses in-depth learning to explore the value of MICE scenarios and uses in-depth learning to study behavior recognition and classification of MICE scenarios [2]. At the same time, the traditional way of video surveillance used in convention and exhibition tourism scene information is costly, easy to miss, inefficient, and relatively less accurate. The convolutional neural network in a deep learning algorithm can simulate the process of human visual information processing, extract the features of self-learning images, improve the accuracy of human behavior, and make the classification effect of convention and exhibition tourism scene more ideal [3].

Currently, the classification of scenes in the domain of computer vision is a research flashpoint. It has firm use-value and implications for the development of the industry. Though, there are few research and related algorithms on mice tourism scene classification. Hence, this research provides a model for behavior recognition using in-depth learning. The proposed model applies the mice tourism scene classification. The scene classification algorithm based on a self-attention generation countermeasure network is constructed. The proposed model collects many visual data and speeds up the rapid development of the field of vision.

The major objectives of the research include a brief description of the convolutional neural network, and then the algorithms based on in-depth learning are mainly described. Behavior recognition technology in static images based on knowledge primitives is used to classify exhibition tourism scenes using a scene classification algorithm based on self-attention generation against a network [4]. Select three public behavior data to verify the validity of behavior recognition. The results show that the expanded data can improve the accuracy and effectively prove that the

algorithm works well in behavior recognition. Simultaneously, the validity of the SAGAN (Self-Attention Generative Adversarial Network) algorithm is verified by comparing the classification results of different models. The result shows that the detection accuracy of this algorithm is as high.

The rest of this paper is organized as follows. In Section 2, related work is provided. In Section 3, behavior recognition based on deep learning approach is discussed. In Section 4, the proposed conventional and exhibition tour scene classification based on in-depth learning approach is discussed. In Section 5, experimental results and discussion are provided about the proposed work to show its effectiveness. Finally, the paper concludes in Section 6.

2. Related Work

There are many scientific research institutes in the world cooperating with universities to study human behavior recognition technology, which uses a computer to analyze video sequences and automatically detect and identify human behavior in video images [5]. Attia et al. expand the traditional neural network (CNN) to the three-dimensional CNN of time information, calculating the spatial and temporal dimensions of video data characteristics [6]. Mathai and other experts use a dual-resolution convolutional neural network to input and process video images into two sets of highly independent data streams, namely, the original resolution and the low resolution, which are alternately composed of regular, convolution, and pooling layers in feature extraction, and fuse the data on the full connection layer, which makes it easier to understand and identify the subsequent signs based on the results obtained [7]. Zahid et al. scholars used human behavior recognition technology to train national team divers and established models to track and identify human behavior [8].

Microsoft Asia Research Institute judges and identifies human gesture motion recognition and uses this result in real-world situations [9]. Experts such as Ma et al. [10] have studied the development of convolutional neural networks from object recognition to behavior recognition and designed and developed a data enhancement mechanism. Zhao and Huang use advanced context cues to build a behavior recognition algorithm based on the R^* CNN framework [11]. Gao et al. use a convolutional neural network to extract unstructured travel text information and establish a travel knowledge map by optimizing the extracted results [12]. Jia t al. used the knowledge map-based recommendation algorithm widely in different fields. It is used to extract the entity-relationship label features within the knowledge map. This method has a low recommendation rate. For this problem, the network embedding method is used to extract the features of the tourism knowledge map.

The semantic features of the map nodes such as tourism and scenic spots are extracted based on the in-depth learning model. Thus, a feature vector of tourists and attractions incorporating different tag semantics is obtained. Ma et al. use multimodal data recognition in online reviews of travel products to analyze valuable data in online reviews and to analyze user-generated optimization methods for online travel products, embedding text vectors and recognition push-by-content based on in-depth and machine learning [13]. Dong and Wang use computer in-depth learning technology to identify more than 30,000 Beijing photos taken by tourists on Flickr and get 103 scenes [14].

3. Behavior Recognition Based on Deep Learning

The convolutional neural network (CNN) is utilized to recognize the behavior in various scenes for the tour classification.

3.1. Convolutional Neural Network. A CNN is a neural network with many hierarchical structures. When using a convolutional neural network for classification and recognition, the two most important parts are classifier and feature extraction [15]. The components of feature extraction include the pooling layer and multiple different convolutional layers. Generally, the classifier contains 1–3 fully connected layers. The convolutional layer (conv) uses many convolutional layers to extract different features on the image. In the process of feature extraction, the local pixels on the input image relate to all convolutional layers to extract the local features of the image, and then the convolution kernel is moved in the order of top to bottom and left to right with a fixed length to extract the global features on the image. Figure 1 shows the convolution operation process, setting 1 as the step size.

Add the offset value of this layer to the convolution result and use the function to activate and obtain the C characteristic diagram on this layer. Equation (1) represents the actual operation.

$$c = f(\mathrm{Conv}(W, u) + b), \tag{1}$$

where f is the excitation function.b is the offset matrix on this layer.W is the filter weight matrix.

The offset matric is used to represent the input image of this layer or the output characteristic diagram of the upper layer. The function of the pooling layer is to preserve all the critical data on the image and reduce the network parameters by reducing the size of the image; at the same time, the network generalization ability and recognition ability are stronger [16]. At present, the most popular pooling methods are maximum pooling and average pooling. Two different pooling operation modes are detailed in Figure 2. Setting a pooled window and an input image size of 2 in this image $\times 2$, 4×4, set the moving step to 2. Maximum pooling is the maximum value in the obscured book image window. Average persistence refers to the average value on the output image that is obscured by the pooled window. After processing, the feature map of the pooled layer extraction can be obtained.

The function of the full connection layer on the classifier is to use the weight matrix to combine the previous features to form an image whole and further select an appropriate classifier to generate classification labels for the original map, to perform the classification task. Softmax is a more widely

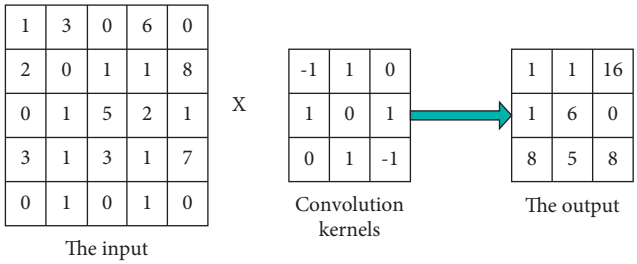

Figure 1: Convolution operational process.

used classifier in deep learning. There are m training samples in the process of performing the k-class behavior classification task. Each sample is an n-dimensional vector, and Dtrain represents the training set, which is represented by

$$D_{\text{train}} = \left\{\left(x^{(i)}, y^{(i)}\right), \ldots, \left(x^{(i)}, y^{(i)}\right)\right\}, \quad i \in [1, m], \tag{2}$$

$x^{(i)}$ are the real labels of the first input sample. $y^{(i)}$ are the first input sample. $y^{(i)} = \{1, 2, \ldots\ldots, k\}$.

The softmax classifier calculates the probability of different behavior types for all input $x^{(i)}$, expressed by

$$P\left(y = j \mid x^{(i)}\right), \quad (j = 1, \ldots\ldots, k). \tag{3}$$

The output value on the softmax classifier is a k-dimensional vector composed of the probability values of different behavior types of $x^{(i)}$ input samples. The prediction vector function is calculated by

$$y(i) = f(x(i)|\theta) = \begin{bmatrix} p\left(y^{(i)} = 1 \mid x^{(i)}, \theta\right) \\ p\left(y^{(i)} = 2 \mid x^{(i)}, \theta\right) \\ \cdots \\ p\left(y^{(i)} = k \mid x^{(i)}, \theta\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \cdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}. \tag{4}$$

Upper form θ for the basic parameters to be learned on the network, the corresponding output vector $y^{(i)}$ can be obtained from the $x^{(i)}$ input samples. There are k values in $y^{(i)}$, each value has a corresponding value interval, and the cumulative total value is 1. In this case, the k-dimensional vector $y^{(i)}$ obtained by the softmax classifier is a probability or score feature on the behavior recognition classification, and each value represents the probability of each row corresponding to the input sample.

3.2. Behavior Recognition Technology in Static Image Based on Knowledge Primitive. Image recognition technology has developed rapidly in recent years and has achieved good results in identifying different objects. However, the performance of human motion recognition is not good. The problem is that the external environment is uncontrollable, and the image collection cannot be better, for example, affected by external factors such as background, light, shadow, and occlusion in the environment. Behavior recognition is an important research direction of computer vision. It has a wide range of applications. It is because the interference of the above factors makes behavior recognition more difficult.

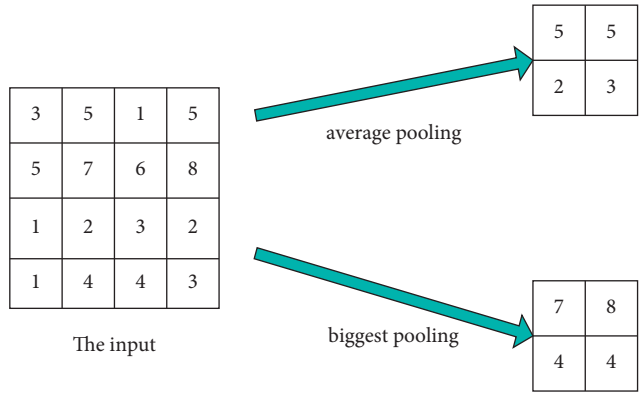

Figure 2: Common pooling processes.

3.2.1. Behavioral Knowledge Framework. This paper first identifies the behavior image that needs to be detected, extracts the knowledge primitive histogram from the image, and then inputs the trained image information into the multi-class linear SVM classifier to predict its behavior category [17]. The purpose of this paper is to develop a behavior framework to complete feature extraction through simple data processing and to code its behavior knowledge. That is, there is no need for more complex feature design processes and label images in the framework. In this framework, the behavior knowledge primitive is considered as a dictionary of behavior posture, $\{I_i\}_{i=1}^{N}$ training image size is $m \times n$, and number N size of each image block is $k1 \times k2$.

3.2.2. Behavioral Knowledge Coding. After obtaining the pooled features of the training image, the k-means algorithm is used and the cluster center is called AKPs. The behavior image is represented by AKPs. The experimental results show that running independently can effectively improve model performance, so the C-class behavior knowledge primitives can be formalized as

$$Q^c = [Q_1, Q_1, \ldots, Q_L], \tag{5}$$

$$\text{also } C \in \{1, 2, \ldots, C\}. \tag{6}$$

3.2.3. Behavior Representation. After the I behavior image is determined, its R_t Rt global feature is extracted, and then the final H_I feature is obtained. The I behavior image is

represented by each element's knowledge of the behavior primitive. That is, the Euclidean distance is used to find the behavior knowledge with the highest degree of similarity. All numbers add up to 1, and the following equations (e.g., equations (7) and (8)) are their expression.

$$H_I = \left[H_I^1, H_I^2, , H_I^C\right], \tag{7}$$

$$H_I^i = \left[\text{dist}\left(R_I(i), Q_j\right)\right]. \tag{8}$$

The i global feature on top image I is represented by $R_I(i)$, and the Euclidean distance is represented by $\text{dist}(\cdot)$ express.

4. Conventional and Exhibition Tour Scene Classification Based on In-Depth Learning

The image of a convention travel scene is used to represent all the images taken in the convention travel environment. The images come from the images of tourist attractions taken by tourists, the images of exhibition products taken by tourists attending the convention, etc. Classification of exhibition tourism scenes means that an image of exhibition tourism is input and output from the model by category labels. Generally, there are three periods to go through. First, the related image data are input into the model, then the feature extraction on the image data is used to train the classifier, and then the classifier is used to predict the image label. That is, the classification of convention and exhibition tourism scenes is to automatically label the aged tourism scene images as different types by computer to achieve the purpose of automatic classification of convention and exhibition tourism scenes. When dealing with the classification of convention and exhibition tourism scenes, this paper uses a scene classification algorithm based on self-attention generation to generate an antagonistic network. The following highlights.

4.1. Scene Classification Algorithm for Generating Antagonistic Networks Based on Self-Attention. The scene classification is achieved by proposing a specific algorithm. The algorithm is responsible for generating antagonistic networks. The antagonistic network generation is based on self-attention. The detailed process of the algorithm proposed is discussed in this section.

4.1.1. Generating the Antagonistic Network. The components of the generated antagonism network include discriminators and generators, which are processed by the antagonism training mode. The generator generates the latest and very close sample distribution based on the actual distribution of the input samples. The discriminator judges its source based on the input samples, and the analysis belongs to the generator generating samples or real samples. The training consists of the following two parts:

(a) The generator continuously captures the distributed true probability data and adds random noise to form a new false data. Optimize a fixed discriminator parameter for the generator parameter, output a high probability value, and minimize $1 - D(G(z))$ within the function.

(b) By discriminator for sample authenticity, a fixed generator parameter is formed by optimizing the parameters in the discriminator, and a high probability value is output to maximize function $D(x)$.

After the above operations, the antagonistic network objective functions are obtained using

$$\min_G \max_D V(D,G) = E_X \sim P_z(z)[\log(1 - D(G(z)))]. \tag{9}$$

4.1.2. Attention Mechanism. The high complexity of the model on the neural network indicates that there is a large amount of information stored, but it will result in an ineffective match between computing power and information overload. The human brain uses the attention mechanism to process various kinds of information. People only focus on the main part of information and ignore other information. Therefore, the overload problem can be better handled by simulating the attention mechanism with limited computational performance.

4.2. Self-Aware Generation of Antagonistic Networks. Image data of exhibition tourism scenes are more redundant than other data, so valuable information about scene classification cannot be obtained. Generating antagonistic network is an unsupervised learning model. The theoretical analysis can meet the requirements of real data distribution. However, the generation of antagonistic networks is a local operation that produces high-pixel space from low-pixel space. It can only process information near the network but cannot compute feature information far away. Experts and scholars have established a self-attention generation antagonism network (SAGAN) model to expand the perception field, which is essentially based on the addition of a new self-attention module to the antagonism network model, as shown in Figure 3. In this paper, the sub-attention module is introduced for two reasons.

First, the model memory of most image features makes it more complex, while the low computing power focuses on the valuable features. In addition, to balance the algorithmic capability with the model complexity, full connectivity can enhance the perception and reduce the computational efficiency of the model. Local join, shared weights, and spooling layers are often added to the optimization, and memory for long-distance features is slightly inadequate. Simulate the attention processing mechanism used by the human brain to process information overload, and add the attention mechanism to the field of artificial intelligence to help the neural network select high-quality information. The SAGAN generator generates new features in various details, and the discriminator can judge the fine features of two images. Add a self-defining module to this problem to

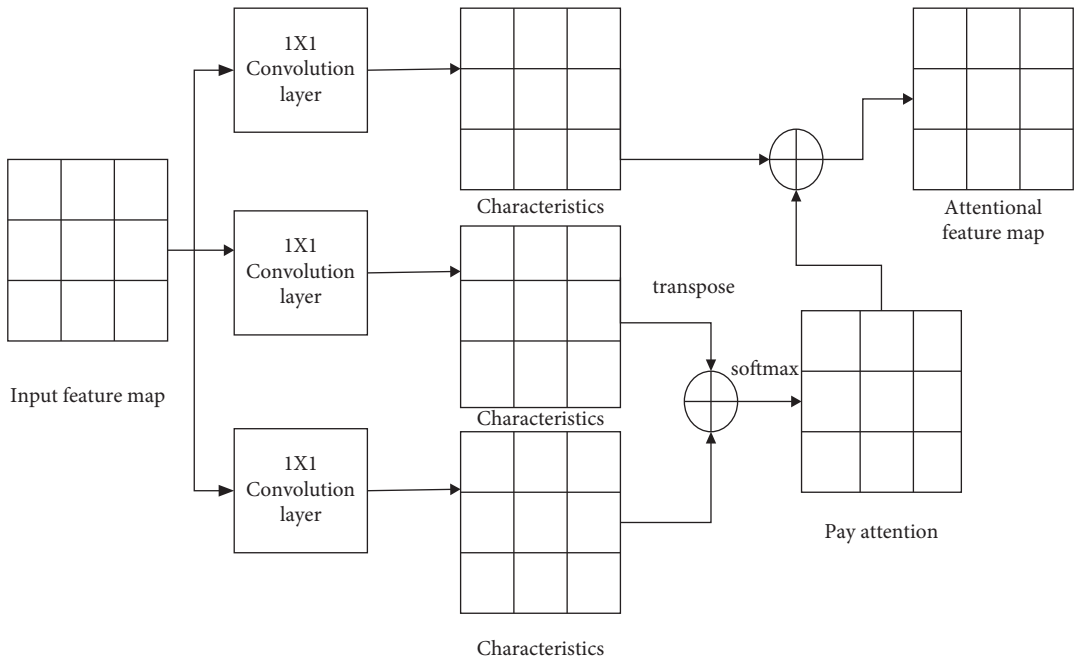

Figure 3: Self-notice module structure diagram.

handle the detailed flaws caused by the source network's inability to capture collection structure problems.

The computational effort introduced in the attention mechanism is small and does not affect the computational efficiency of the algorithm. You need to go to three 1×1 convolution kernel convolutional layer input feature $x = R^{c \times N}$ on the hidden layer, and through semantic transformation, the following three functions can be obtained, that is, $g(x) = w_g \cdot x$; $f(x) = w_f \cdot x$; $h(x) = w_h \cdot x$, and the learned privilege matrix can be expressed as $w_g \in R^{\bar{c} \times c}$; $w_f \in R^{\bar{c} \times c}$; $w_h \in R^{\bar{c} \times c}$.

After transposing the reconstructed $g(x)$ and $f(x)$ eigenvalues and multiplying the matrix, the eigenvalue matrix obtained is the attentional map after the function of softmax. Equation (10) is the spatial correlations of two different eigenvalues.

$$\beta_{j,i} = \frac{\exp\left(f\left(x_i\right)^T g\left(x_j\right)\right.}{\sum_{i=1}^{N} \exp\left(f x_i\right)^T g\left(x_j\right)}. \tag{10}$$

Matrix multiplication between $h(x)$ and attention map to obtain attention character map from a convolution structure of 1×1 is

$$o = \left(o_1, o_2, \ldots, o_j, \ldots, o_N\right) \in R^{C \times N}. \tag{11}$$

Global spatial information is represented as

$$o_j = \sum_{i=1}^{N} \beta_{j,i} h. \tag{12}$$

Figure 4 shows a flowchart of the basic framework for self-attention generation of antagonistic networks. In addition, by integrating local and global spatial information, that is, by multiplying the output on the attention layer by the coefficient y, y_i can be obtained using equation (13) further if the x feature map is used.

$$y_i = \gamma o_i + x_i. \tag{13}$$

Initialize first γ, a value of 0 focuses on relatively simple neighborhood information and then adds attention to more complex distance features to make feature extraction more effective. The autonomous module is added to the discriminator and generator as shown in Figure 4.

SAGAN objective function uses alternating training to minimize antagonistic loss using

$$\begin{aligned} L_D = & -E_{(x,y)\sim \text{pdata}}[\min(0, -1 + D(x, y))] \\ & - E_{z \sim pz}[\min(0, -1 - D(G(z), y))], \end{aligned} \tag{14}$$

$$L_G = -E_{z \sim p, y \sim \text{pdata}} D(G(z), y). \tag{15}$$

In this paper, we use the self-definition method to better balance the hyperparameter sensitivity of GAN. SAGAN has a smaller number of parameters than full connection, which balances the reduction of parameters with the enhancement of the receptive field.

5. Analysis of Classification Results of Exhibition Tourism Scenes

The experimental analysis of the proposed classification model is presented in this section. The results are elaborated in the context of the of exhibition tourism scenes.

5.1. Behavior Recognition Analysis Results. The proposed model is based on in-depth learning to study the problem of behavior recognition and categorization of exhibition

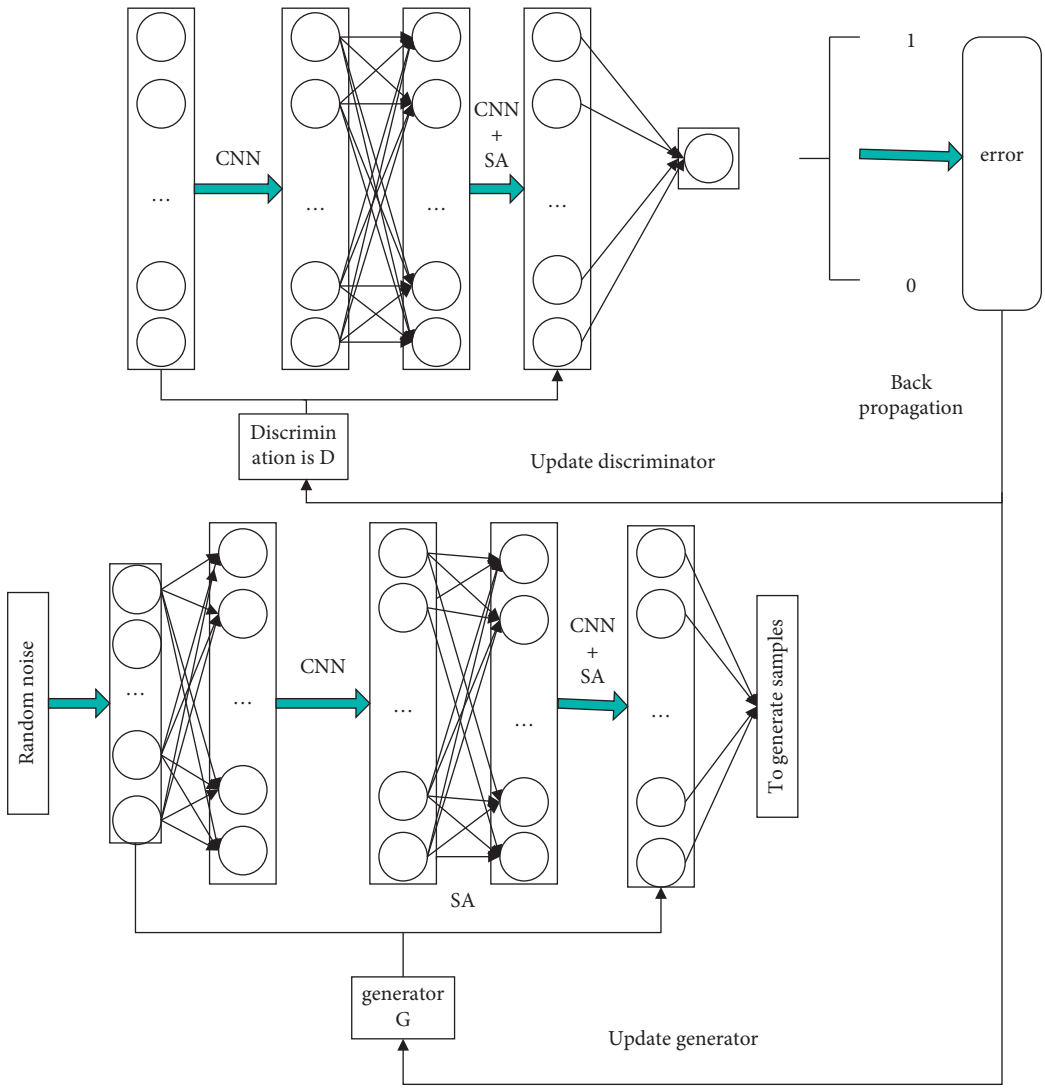

Figure 4: Flowchart of the basic framework for self-attention generation of antagonistic networks.

scenarios, in which the results of behavioral analysis are described in detail. Two strategies are used to predict the recognition image class label after the neural network has been trained. Initially, the behavior label is predicted using the softmax classifier value. Or a multi-class linear support vector machine can be obtained by using 4096-dimensional training features. The strategy considers the neural network as a classifier and extractor with feature extractor. The basic model utilized here is VGG-16, which is used for completion of the neural network architecture. On the convolutional layer, the convolution kernel is 3×3. To complete the maximum pooling process, use the ReLU excitation function at the implied layer and the spatial sliding window with a step of 2 and a size of 3×3 to lower the spatial dimension on the feature map. Set the dropout value probability to 0.5, and apply the dropout operation to the first two completely connected layers. The number of "FC8 Action" neurons is equal to the number of behavior categories in the dataset, but "FC8 Hint" has a dimension of 150. At this point, all the networks on the "FC8_Hint" layer can be discarded to assess and verify the validity of the model. The test process first determines the images that need to be tested and then expands and clips the same data to form a test picture block. Next, the picture is quickly input to the neural network [18] after RGB dithering and random horizontal flipping. Prediction using two recognition methods starts with a score prediction on all image blocks, and as a prediction score for another complete image, the class labels corresponding to the highest score are output. To obtain 4096-dimensional features, extract 4096-dimensional features from all picture blocks and then average all image block features. Finally, to predict behavior class labels, a 4096-dimensional feature is entered into the learned SVM classifier.

This research chooses three public behavior datasets to test the validity of this algorithm, mainly PPMI, Stanford-40, and PASCAL_VOC2012. The average accuracy (mAP) of the three training datasets is calculated according to the evaluation principle of this method by comparing several algorithms that currently work well. The performance of different experimental configurations is listed in Table 1. A cross sign indicates that they are not selected, and a check sign indicates that they are selected.

Table 1: Performance of different experimental configurations.

Methods	AugData	Hint	Deep	Com	SVM	Softmax	mAP
NAug	×	×	×	×	×	√	71.58
NAugDS	×	×	√	×	√	×	75.28
NAugCS	×	×	×	√	√	×	75.32
Aug	√	×	×	×	×	√	76.46
AugDS	√	×	√	×	√	×	77.42
AugCS	√	×	×	√	√	×	77.58
AugH	√	√	×	×	×	√	78.33
AugHS	√	√	×	×	√	×	80.7

This paper evaluates the impact of two ways of assisting hint task and data enhancement on Stanford-40 datasets. Four different ways of assessing the effect are listed in Table 1. "NoAug" means that no ancillary tasks or database extension techniques were used to adjust the model and make the first identification strategy prediction. "NoAugCS" and "NoAugDS" strategy use the same prediction method, the difference is reflected in one using 4096-dimensional features and the other using mixed features. "AugD S" trains a neural network using ancillary tasks and data expansion. "Deep" is 4096-dimensional feature, and "Com" is a brand-new feature which combines 04096-dimensional feature of "Deep" with PAVs feature. The analysis of the results shows that the recognition accuracy can be improved by 2% through data expansion, which fully proves its validity. The accuracy of "Com" is 0.17% higher than Deep. If PAV is used for assistant task design and the results of AugD S and AugHS are compared and analyzed, the accuracy of assistant task can be improved by about 3%. Figure 5 shows the impact of data and size on performance.

5.2. Classification Results of Exhibition Tour Scenes. In this paper, two strategies are used to collect exhibition tourism scene data, extracting specific address information to centralize the management of exhibition and tourism scene photos in the same location [19]. Street panoramic parameters and basic values are listed in Table 2.

This paper chooses the data interface of street panorama as the data source when classifying convention and exhibition scenarios. The panoramic street pictures in Table 2 have a large coverage area. By analyzing the address parameters in the panoramic image, a unique parameter can be obtained, and data pictures can be taken from multiple perspectives which are close to each other. By filtering the data and adjusting the data format, as in-depth learning has a high requirement for training models, the original data can be expanded by conventional data expansion methods such as rotation, translation, and scaling. After the expansion, the pictures can be increased to 2,700,000. Pre-data preparation process should be divided into different datasets in 7 : 1 : 2 ratio, corresponding to test set, training set, and validation set.

In the experiment, a 16 GB memory GPU was selected to compute the SAGAN-based image classification of tourism scenes, and the image classification effect was calculated using this method. Set the model learning rate and batch

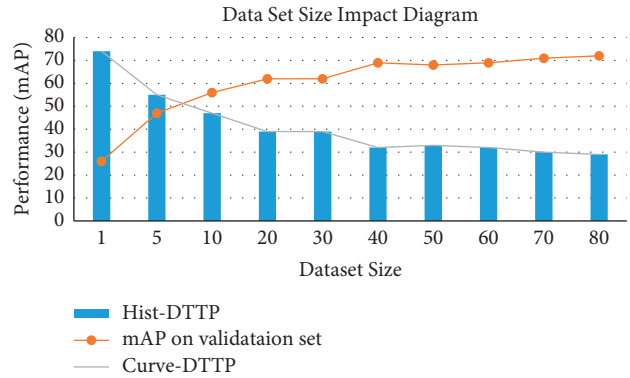

Figure 5: Dataset size impact diagram.

Table 2: Street panoramic parameters and basic values.

Parameter	Instructions	Value of the sample
Pano	Street scene ID, unique value	Pano = 10011234485
Heading	Yaw angle; the value ranges from 0 to 360	Heading = 90
Pitch	Pitch angle: 0° is level, 90° is directly below, −90° is directly above, level by default	Pitch = 90
Size	Image size	Size = 640 × 480

number of 0.0003 and 128, respectively. Here, the IS value is introduced into the generator capability test, which is a data index to evaluate the model performance based on the relationship between the variety and quality of the generated image, which can more intuitively show the degree of fit between the real sample and the generated sample. Show model generated sample IS values and true sample IS values in Table 3.

The image classification experiments of exhibition image scenes are carried out using proposed classification model based on the existing models. The results are shown in Table 4 to judge the ability of the SAGAN model to classify this kind of image. According to these data, the current neural network model is a general model, which can classify the data more accurately, and the accuracy of classification results is higher than 97.5%. The classification results from multiple models are listed in Table 3.

The classification of the scenes is observed with training data from 10% to 80% in different scales using the prosed model. The results are shown in Figure 6. The current CNN

Table 3: Sample IS values generated by each model.

Real sample	SAGAN	DCGAN	VAE
4.53	3.51	3.12	2.95

Table 4: Classification results for different models.

Model	Accuracy (%)	The elapsed time
AlexNet	98.13	5263
VGG-VD19	97.52	14943
SSGAN	98.34	8421
SAGAN	99.12	5296

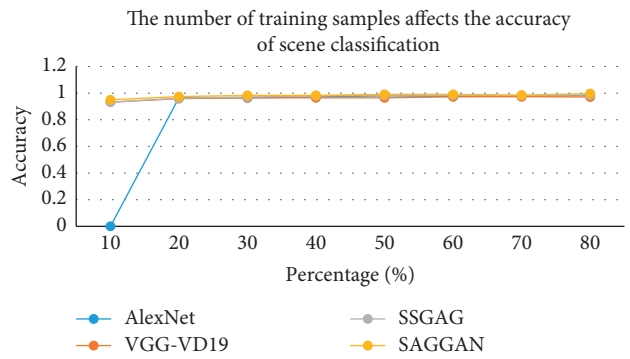

Figure 6: The number of training samples affects the accuracy of scene classification results.

image classification results are good, and there is a lack of accurate classification from the analysis of exhibition tourism. The reason for this is that it is not possible to get a more accurate classification from a certain feature because the scene features are not significant. Secondly, there are not enough samples of the same type. At the same time, the current model cannot get accurate scene content based on the lack of enough samples. Figure 6 shows how the number of training samples affects the accuracy of scene classification.

6. Conclusions

With the rapid development of science and technology and the comprehensive popularity of mobile devices such as cameras and smartphones, people are rapidly entering the era of intelligence. As the main media for people's daily communication, the network produces a huge amount of visual data, which involves a variety of video and image information. With the rapid development of computer technology and network technology, the exhibition tourism industry has become an important industry to promote the economic development of the region. However, with the increase in exhibition tourism data, people cannot effectively classify exhibition tourism scenes. For this problem, this paper uses an in-depth learning algorithm to study behavior recognition and classification of exhibition tourism scenes. First, the neural network is briefly introduced, the convolutional neural operation process is described, and the behavior recognition technology in static images based on neuroprimitives is described in detail. Then, it uses the scene classification method based on self-attention generation to classify the exhibition tourism scenes and compares the performance of different configurations by experiment to test the effect of behavior recognition. Using two strategies to extract data from exhibition tourism scenes, this paper uses the scene classification algorithm based on self-attention generation to generate an antagonistic network with the highest accuracy, reaching 99.12%. The results show that the algorithm can accurately classify the convention and exhibition tourism scenes.

Data Availability

All the data are included in this paper.

Conflicts of Interest

No conflict of interest exists for publication of this paper.

References

[1] Z. Ma, "Human action recognition in smart cultural tourism based on fusion techniques of virtual reality and SOM neural network," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 3495203, 12 pages, 2021.

[2] S. Safaeva, D. Adilova, and D. Adilova, "Mice tourism: opportunities, priorities, problems, prospects," *American Journal of Applied Sciences*, vol. 02, no. 11, pp. 116–121, 2020.

[3] D. Mohammad, I. Aljarrah, and M. Jarrah, "Searching surveillance video contents using convolutional neural network," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, p. 1656, 2021.

[4] J. Zhang, T. Wu, and Z. Fan, "Research on precision marketing model of tourism industry based on user's mobile behavior trajectory," *Mobile Information Systems*, vol. 2019, Article ID 6560848, 14 pages, 2019.

[5] Y.-H. Byeon, D. Kim, J. Lee, and K.-C. Kwak, "Body and hand-object ROI-based behavior recognition using deep learning," *Sensors*, vol. 21, no. 5, p. 1838, 2021.

[6] A. Attia, Z. Akhtar, and Y. Chahir, "Feature-level fusion of major and minor dorsal finger knuckle patterns for person authentication," *Signal, Image and Video Processing*, vol. 15, no. 4, pp. 851–859, 2021.

[7] A. Mathai, L. Mengdi, S. Lau, N. Guo, Q. Guo, and X. Wang, "Transparent object reconstruction based on compressive sensing and super-resolution convolutional neural network," *Photonic Sensors*, vol. 12, no. 4, 2022.

[8] F. B. Zahid, Z. C. Ong, S. Y. Khoo, and M. F. Mohd Salleh, "Inertial sensor based human behavior recognition in modal testing using machine learning approach," *Measurement Science and Technology*, vol. 32, no. 11, Article ID 115905, 2021.

[9] B. Javidi, F. Pla, J. M. Sotoca et al., "Fundamentals of automated human gesture recognition using 3D integral imaging: a tutorial," *Advances in Optics and Photonics*, vol. 12, no. 4, p. 1237, 2020.

[10] T. Ma, Y. X. Hu, and X. H. Zhang, "Deep reinforcement learning based coflow scheduling in data center networks," *Acta Electronica Sinica*, vol. 46, no. 7, pp. 1617–1624, 2018.

[11] X. H. Zhao and C. L. Huang, "Research on identification algorithm of mine person's violation behavior based on

kinect," *Journal of Hunan University*, vol. 47, no. 4, pp. 92–98, 2020.

[12] J. L. Gao, P. Y. Qiu, L. Yu, Z. C. Huang, and F. Lu, "An interpretable attraction recommendation method based on knowledge graph," *Science in China(Information Sciences)*, vol. 50, no. 7, pp. 1055–1068, 2020.

[13] C. Ma, G. Li, S. J. Chen, J. Mao, and Q. Zhang, "Research on usefulness recognition of tourism online reviews based on multimodal data semantic fusion," *Journal of the China Society for Scientific and Technical Information*, vol. 39, no. 2, pp. 199–207, 2020.

[14] Y. Dong and A. Q. Wang, "Image recognition technology based on deep learning," *Modern Industrial Economy and Informationization*, vol. 11, no. 6, pp. 154-155, 2021.

[15] D. X. Yu, B. M. Zhang, C. Zhao, H. T. Guo, and J. Lu, "Scene classification of remote sensing image using ensemble convolutional neural network," *Journal of Remote Sensing*, vol. 24, no. 6, pp. 717–727, 2020.

[16] Y. Zeng, Y. L. Chen, and X. D. Cai, "Face recognition algorithm for the deep hash combined with global and local pooling," *Journal of Xidian University*, vol. 45, no. 5, pp. 163–169, 2018.

[17] L. M. Zhao, B. Zhu, T. Bai, and Y. Z. He, "Human behavior recognition based on image recognition technology," *Industrial Control Computer*, vol. 34, no. 2, pp. 107-108, 2021.

[18] H. J. Yu and Y. C. Fang, "A robot scene recognition method based on improved autonomous developmental network," *Acta Automatica Sinica*, vol. 47, no. 7, pp. 1530–1538, 2021.

[19] Y. F. Yang and W. L. Chen, "Joint model for entity alias extraction in tourism domain," *Journal of Chinese Information Processing*, vol. 34, no. 6, pp. 55–63, 2020.

[20] W. Y. Liu, Z. C. Guo, and D. D. Guo, "Text aspect-level fine-grained sentiment classification based on deep learning in travel scenarios," *Chinese High Technology Letters*, vol. 32, no. 1, pp. 22–32, 2022.